From Transitive Inference To Exhaustive Search: Towards Self-Regulating Models of Developmental Processes

Benedict St Johnston
Centre for Cognitive Science
Edinburgh University

Thesis submitted for the degree of PhD

November 1994

Abstract

This thesis is sympathetic to Piaget's view of development as a dynamic interactive process between an autonomous agent and its environment allowing cognitive growth through self-regulation and resulting in an extended behavioural repertoire. This thesis explores developmental issues through computer modelling.

Piaget thought transitive inferences relied on logic, which he considered the basis of rationality, and as such marked the end-point of cognitive development. More recently, empirical studies have shown that the ability to make transitive inferences does not rely on logic, and is close to the lower ontological bounds of the system, requiring us to reassess our theories of rationality, ontology, and ontogeny.

Using Piagetian principles, it is argued that transitive inference is likely to be primitive, and should be applied as a default assumption for pragmatic reasons but that this assumption must be defeasible.

Basic decision-making models are produced, based on Sutton and Barto's self-supervised learning models. They show how easy it is to generate a simple order over a set of binary choices leading to a strong transitive bias. Elaborating the representation in the model with units that learn stimulus-stimulus relationships, the model can also learn a circular relationship and generalise appropriately. Thus, transitivity is captured as a default yet defeasible assumption giving the agent the ability to exploit its environment as much as possible whilst still remaining adaptable. The model also shows similar scope to subjects in that its performance on triadic choices is worse than on binary ones. Unlike some subjects, the model cannot repair its own performance.

Seriation is a more transparent ordering skill which develops later. Empirical work determining the basis of seriation skills showed that the important development is a data-reducing strategy, based on picking up redundant information in the task, that reduces seriation to a simple search task. This, taken with the limitations in transitivity, suggested that what develops is not new skills, but an extension of the scope of existing skills through the use of simplifying strategies. Thus a new ontology is emerging. The driving force behind development now seems more likely to be resource management than truth preservation, as Piaget supposed.

The exhaustive search paradigm was designed to investigate how resource management might lead subjects to constrain their search strategies in order to reduce memory load leading to greater efficiency. This task is analysed to provide measures of success against which subjects performance can be evaluated and to assess the difficulty of some of the problems that the subjects must overcome to perform well. Successful children maximally constrain their behaviour showing strong adjacency and vector constraints. Monkeys performance improves indicating self-regulation, but their searches are not constrained to the same degree.

An exploratory modelling strategy is used to determine the minimal set of resources required to solve the task and the different constraints that might lead to strategic development. The modelling indicates that not only must the subject effectively reduce the number of items considered at each point during the search, (through vector and adjacency constraints for example), but also reduce the total production length through chunking.

The importance of chunking and classification is discussed in relation to the new characterisation of development.

In accordance with the requirements of the University of Edinburgh, regulation 3.4.7, this thesis has been composed by myself and the work presented herein, except where explicitly stated otherwise, is my own.

Benedict St Johnston

Acknowledgements

Without the intellectual and moral support of my supervisors Keith Stenning and Brendan McGonigle, this thesis would not have happened.

Steve Finch helped a great deal in the development of the probability model of exhaustive search. Without Carlo de Lillo and Tony Dickinson, I would not have had any experimental data on the exhaustive search task to model. They were also part of the research group including Brendan and Margaret Chalmers, that kept me going.

My parents supported me spiritually and financially throughout, and I hope that their extreme scepticism at what I was trying to do has now been proved wrong.

Many thanks to the Browns for listening interestedly, when I tried to explain what I was doing and providing countless meals to build up my strength.

Without Brendan's brutal and justified criticism of the first draft of this thesis, it would have been much inferior.

Lastly, thanks to Morag, who suffered more than anyone during these long years. She was always there pushing me on, and generally irritating me so much that I had to work.

To the people mentioned above I am eternally grateful -- overall I thoroughly enjoyed myself.

Table of Contents

Overview of the Thesis

Chapter 1 begins by setting out what cognitive development is, why is it interesting and what sorts of problems need to be solved. A brief survey of alternative theories of development set the scene for the philosophy behind Piaget's theory. Piaget believed development to be a creative adaptive process through which cognitive structures are built allowing the developing agent to make better sense of the world and so become more competent. Each construction builds on what already exists and is achieved through interaction with the world. Broadly speaking, this is the view of development accepted in this thesis and it has two important implications. Firstly, we must understand development from the bottom up. Secondly, such an interactive process needs something to drive it. Piaget claimed that we begin life with little more than reflexes and develop through a number of stages progressively allowing more concrete practical reasoning and finally leading to abstract logical reasoning. He believed this process was driven by truth preservation and that the ability to make transitive inferences marks the end-point of development. This view is rejected on theoretical grounds, (truth preservation cannot lead to the development of more and more complex logics), and on empirical grounds, (transitive inference seems to be a low level developmentally primitive ability). Thus, we need to determine a new developmental trajectory and a new driving force. Given the new status of transitive inference, this seems an excellent place to start.

It is then argued that previous research on transitive inference has ignored the fact that it is usually a lot harder to determine whether a relation is transitive or not, than to make the appropriate inferences afterwards. However, given that an agent must often deal with indeterminate relations, assuming transitivity is a pragmatic thing to do, but this assumption must be defeasible in the light of new information. This new characterisation of transitive inference problems presents a challenge to any model of the skill.

Chapter 2 begins by explaining the purpose of modelling and computer simu-

lation, and details some of the problems associated with this methodology. Some existing models of transitive inference are then reviewed in the light of the new characterisation presented in chapter 1.

Chapter 3 covers the modelling of transitive inference. Using Sutton and Barto's Temporal Difference Learning Framework it is shown how a weak stochastic transitive bias is inherent in this type of model. It is then demonstrated just how easy it is to magnify this inherent bias. All that is required is to elaborate the input units with some other units. Any type of inputs at all will do – even random noise. This clearly vindicates the view taken, that getting a model to produce a strong transitive bias is not impressive in itself and that models should also be able to capture intransitive reasoning. The models can learn a circular relationship but in order to be able to generalise usefully special units which learn contiguity relations between stimuli must be added to the input layer. These contiguity units also allow the model to assume transitivity as a default *and* this assumption is defeasible.

There were two important experimental effects that the models could not capture. Most importantly, when subjects first learn a 5-term transitive series and are then given post-tests of 3 items, their initial performance drops but can recover spontaneously without differential reinforcement. The model with contiguity units showed the initial drop but could not spontaneously recover. Thus there seems to be some internal self-regulation in the subjects that is not incorporated into the model. Secondly, there is evidence that subjects learn the premises faster when they they are presented in a serial as opposed to a random order. None of the models showed this naturally.

The transitive inference task is not a serial task, nor does it provide much insight into internal self-regulation. A new task is needed to allow the modelling to be constrained enough to tackle these phenomena. At this point the work becomes very closely entwined with the empirical work of McGonigle et al whose data had been the main source for the modelling of transitive inference.

Chapter 4 details the history of this empirical programme starting from the

transitive inference experiments, and leading to the rationale behind the exhaustive search task.

Once transitive choice is understood as a low-level task then the fact that subjects cannot seriate those same items becomes understandable. This raises the question, however, of what exactly it is that develops which allows subjects to seriate. The drop in performance on triads may provide the clue. Subjects can cope perfectly with binary comparisons but if the number of items is increased then these low-level abilities begin to falter. This suggests that what develops is some means of reducing tasks outwith the scope of primitive abilities, to more manageable tasks within the existing scope of the agent. This would also explain why five year old children can seriate five items in a row but not ten.

The classical seriation task devised by Piaget, does not allow the observation of what it is that develops, because it confounds many different skills together. For this reason, McGonigle and Chalmers' empirical programme decomposed the seriation task, and the results suggest that ordinal abilities are not the basis of successful seriation. In fact what seems to underlie success is the ability to reduce the strain on working memory by treating the task as a search task and using redundant information to constrain the search as maximally as possible.

So, the picture of development emerging is one in which we start off with a number of basic abilities that are limited in scope, that the scope of these abilities is extended through the development of data-reducing strategies, and that the driving force behind these developments is resource management.

The seriation task is not a self-regulating task; subjects are driven through a particular serial production and we only find out if this skill is currently within their scope or not. To understand how data-reducing strategies develop through self-regulation driven by resource management we need a new task were the experimenter can manipulate the amount of cognitive strain and the subjects have more freedom to self-regulate their own performance. This is the rationale behind the exhaustive search task where subjects only get rewarded once they have touched every stimulus presented to them on the screen. The subjects have the freedom to

touch the stimuli in any order they wish and they can also touch a stimulus more than once.

Modelling the exhaustive search task presents new problems from the transitive inference modelling. With the latter, all the data had been collected, analysed and characterised before the modelling began making it very clear exactly what is was that the modelling should achieve. With the exhaustive search task, however, the empirical programme was concurrent with the modelling. We did not know how well the subjects would perform and we did not yet have a means of properly measuring their performance. The advantage of modelling in this situation is that the relationship between the modelling and the empirical work becomes a symbiotic one – the data from the empirical programme informs the modelling but also, the modelling helps interpret the data and guide how the empirical programme should progress.

The long-term ambition of the modelling was to capture self-regulation driven by resource management, leading to data-reducing strategies that manifested themselves in constrained search paths. The first step, however, must be to determine what minimal set of cognitive resources are required to achieve competence on the task – you cannot understand resource management without having some idea of the resources available. The modelling strategy, therefore, was to increment a baseline model with additional resources until these proved sufficient for the task at hand.

Chapter 5 begins by describing the experimental procedure for the exhaustive search task. It then details a random model of the task which is necessary for the evaluation of the subjects' and the models' performance. Some possible strategies for the task are analysed, highlighting the difficulties facing the subjects and models. Finally, the initial characterisation of the subjects' performance used to direct the first stage of the modelling is given.

The results from testing children show clearly that if the subject does not constrain their searches with a spatial strategy they forget what they have touched leading to highly inefficient searches. There is a strong correlation between age and performance showing that the spatial strategy is indeed a developmental strategy.

The monkeys provide us with the opportunity to watch spatial strategies develop through exposure to the task.

Chapter 6 presents the modelling of the exhaustive search task. The first model has no memory for where touches have been made so cannot possibly solve the task. Its purpose is to act as an ontological baseline. The next models show how increasing the amount of memory improves the performance of the model but only upto a ceiling of five or six items to be searched. The performance of the models is analysed through decomposing the policy into component strategies utilising starting points, aversion to repeating, and preferred transitions. The similarities and differences between the approach to search taken here and that used in Artificial Intelligence is discussed.

A *reductio ad absurdum* proof shows how a psychologically plausible implementation of working memory cannot possibly lead to optimal performance. Changing the implementation of working memory to overcome this problem actually leads to a deterioration in performance. Further analysis of what the subjects are doing, guides the implementation of additional constraints such as adjacency. None of the additions improves the performance of the model significantly. The reason for this comes from an analysis of how constraints in the search path might become stabilised. It turns out that forwards errors lead to the destabilisation of any such constraints. It is argued that the only way to overcome the problem of forwards errors is to chunk actions together in a more permanent way, such that the problem never becomes serious enough to effect performance significantly. Evidence for the subjects chunking is discussed.

A chunking ability would make the model as competent as the subjects, but the task does not constrain the possible chunking mechanisms sufficiently to make the modelling worthwhile, and the modelling framework also makes it difficult to add such an ability with any ease.

Chapter 7 summarises the results of the thesis and discusses some of the implications for future work, especially in regards to chunking.

Chapter 1

A Perspective on Development and Transitive Inference

1.1 Cognitive Development

1.1.1 What is Cognitive Development?

Cognitive development refers to changes over time in the way we think. Although developmental theories include many nondevelopmental theoretical constructs, such as attention and mental representation, they diverge from nondevelopmental theories by concentrating on changes over time in these constructs.

For example, a child of five cannot arrange ten blocks in a row according to size, and furthermore cannot even appreciate their own inability to do so, (Piaget and Inhelder, 1968). They go through a short period where they know they cannot do the task but are helpless to do anything about it, but by the age of seven they do the task without any trouble at all.

What has changed? How has it changed? And why does it change? These are questions of cognitive development.

1.1.2 Why Should We Study Development?

One might think that it is hard enough to try and determine how we think at one specific point in time, without introducing the extra complexity of developmental change, but it has become increasingly apparent that to understand fundamental

components of thought, such as problem solving for example, we must consider not only innate abilities but an individual's experiental history. This, of course, raises huge problems for the researcher forcing the development of new experimental techniques and the construction of different sorts of model, focussing more and more on the trajectory to expertise in addition to the final competence achieved.

The greater the role of experience in determining the way we think, no matter whether it merely helps calibrate some innate cognitive device or plays a more active role, the greater the importance of a developmental theory.

1.1.3 To Develop or Not to Develop?

Advantages

The "higher" up the phylogenetic scale an organism is, the greater, it seems, the ratio of developmental change over the mere maturation of innate abilities. This casual observation underpins the accepted idea that the advantage that development conveys is an increase in adaptivity. If every ability were static and innate then an organism might be well adapted to one particular niche but would be vulnerable to every environmental change. If on the other hand these cognitive abilities can be moulded by the niche in which the agent happens to find itself then the agent should be more robust to environmental change.

Another advantage of development over the innate specification of behaviour, is that less must be encoded in the genes. For although the environment can be highly variable, it is also highly structured. If an organism can use this external structure to calibrate, organise or even design its own internal cognitive structures then less must be "built-in" from the start.

This is particularly important when you consider that the problems facing an organism will change thoughout its lifespan. For example, merely a change in size will affect its relationship to preditors, prey and competitors, as well as the inanimate world such as whether a branch will support its weight. Obviously,

some changes are innate, such as puberty, but a general ability to adapt reduces the need for these more "costly" timetabled maturations.

So development can be seen as both a way of economising on the genetic information required and a means of allowing adaptability and therefore robustness to the changes in relationship with the environment.

Disadvantages

If development is so useful, then why don't "lower" organisms use it? The answer is that development comes at a great cost. In fact, if we didn't observe it happening all the time we might well have considered it impossible. Some of the problems of development are outlined below, but new problems are coming to light all the time suggesting that we are uncovering the tip of an iceberg.

Developmental growth implies change but an agent must be able to deal with any tasks the world throws at it, at all times during its life. If it fails to deal with the world then it dies. Thus an organism must have a minimal set of competences from the very beginning. In formal systems theory these competences constitute the lower ontological bounds of the system (see Alvarez de Lorenzana and Ward, 1987; LeGare, 1987). These lower ontological bounds then define the ontological space through which the developing system can travel. In other words, development does not happen in a vacuum but is a process that must build on what already exists. Crucially, the lower ontological bounds highly constrain the developmental changes that work. Assuming that these lower ontological bounds have been selected for over countless generations, they constitute at least the minimal set of competences for initial survival. Any developmental change that destroyed one of these minimal competences would be lethal. Thus development cannot merely be a process of trial and error but must be a highly controlled trajectory through a rigidly defined ontological space.

The more complex a system becomes, the more things that can go wrong with that system, but for development to work, requires not only a considerable amount of complexity to begin with, but also a solution to the difficult problem of dealing

with ever increasing complexity. This problem is well illustrated by a problem with semantic networks dubbed the "paradox of interference", (Smith, Adams and Schorr 1978). The more knowledge that is encoded into the semantic network the greater the number of links between nodes. The paradox arises because rather than improving the performance of the network, the increased knowledge leads to interference resulting in longer response times and less accuracy. Somehow, human experts get round this problem, knowing more *and* being more accurate and faster in their responses. How this is possible is still largely unknown. The integration of facts only mildly and temporarily alleviates the problem (Reder and Anderson 1980).

Thus, systems growth, generally speaking, seems to be inherently bad. What this suggests is that in order for cognitive growth to occur at all requires a sophisticated and intelligent metacognitive processes that deal with complexity itself. These processes must not only vet each incremental change but also constantly reduce the overall complexity by removing previous increments that have become redundant. Returning to field of knowledge representation, this idea is captured by the adage: the prerequisite for a good memory is knowing what to forget.

So cognitive development cannot be a simple incremental process. Trajectories through the ontological space will be highly nonmonotonic but, of course, the trajectory cannot leave this space.

No creature devotes so much time and energy to cognitive development than *homo sapiens*. How we achieve this, whilst avoiding the pitfalls, is still miraculous. This thesis constitutes just one small step towards the demystification of this most creative area of cognition.

1.1.4 Methodological Problems

Whilst development is difficult for the developing organism, the problems facing the researcher trying to understand these processes reach new heights. We do not know how a new-born child thinks, but every developmental change must build onto this base. Another obstacle is that the problems of development suggest a

highly sophisticated cognitive architecture but our observations of a developing system cannot cross the bottleneck of behaviour. Knowing that a child cannot solve a task at one age yet can solve it later does not give any information on how or why the developmental change took place, nor what the development actually is. This coupled with the ethical limitations of experimenting on children, means that theorists are not well enough constrained or guided by empirical data, leading to vague and unfounded theories of the mechanisms of change. This thesis attempts to overcome this through two different ways. Firstly, it relies on data from an imaginative research programme in which monkeys and children are used to study developmental processes in detail so the theories can be as heavily constrained by data as possible (see for example, McGonigle and Chalmers 1986, McGonigle et al 1993). Secondly, the theories are instantiated as computer simulations, which constrains the theorising in a useful way and forces every assumption to be explicitly stated (even if only in a programming language). Computer simulations can be considered as self-testing theories. We know immediately we run them whether they are adequate. Of course, adequacy is not the only criterion a theory should meet, but in a field where so much is unknown it is extremely useful to get immediate feedback on a minimal requirement of a theory.

1.2 Alternative Theories of Development

There are many different theories of development. The one espoused in this thesis might be labelled as a synthesis of Piaget's theory and principles from information-processing and engineering. Before coming to Piaget, however, it is worthwhile to survey some of the alternative theories.

1.2.1 Freud

No survey of theories of development could leave out Freud. Freud considered the mind to be made up of different competing components. The **id** is the seat of innate desires or as Freud(1964) put it, the id is the "dark, inaccessible part of

our personality ... a chaos, a cauldron full of seething excitations". The id wants immediate satisfaction according to the **pleasure principle**. The id operates throughout our lifespan but after early infancy, it is joined by the **ego** and the **superego**. The ego is the mind's avenue to the world. The ego is rational, organised and logical, and its activities include perception, logical thought, problem-solving and memory. The id still demands gratification, but the ego allows more considered actions for achieving this gratification. The superego is the last to develop and does not arise until children have resolved their Oedipus complex. The superego is the conscience – it opposes both the id and the ego, in rewarding punishing and making demands.

Freud made two strong claims about development. Firstly, that the first few years of life are the most important – a claim that is now uncontroversial but had not been taken seriously before. Secondly, that development involves psychosexual stages. Each stage is named after a the body part around which drives are centred. If the child develops "normally", they will pass through the oral, the anal, the phallic and finally the genital phase. The details of each stage are not so important to us here as the relationship between stages. Although Piaget, (whose theory will be discussed shortly), also strongly believed in separate identifiable stages of development, he believed that each successive stage built upon and subsumed the previous ones, and that a child would not develop to the next stage until the previous one had been completed. Freud's stages, on the other hand, are completely independent from each other. The transition from one stage to the next is a purely maturational process which will occur whether the child is ready or not, and constitutes a radically different set of drives. Most of Freud's evidence for these stages comes from his explanation of neuroses as the result of a child moving onto the next stage before being ready.

Freud made it very difficult for anyone to criticise his theory. He claimed that without many years of psychoanalysis no-one was qualified to judge in the first place, and anyone who dared to suggest that he overemphasized the role of the infant's sexuality was accused of being sexually repressed themselves. Despite this Freud's methodology, relying as it did exclusively on psychoanalytical techniques

such as dream interpretation and free association, has been totally discredited and no hard evidence can be found in favour of his theory. No one can dispute, however, the enormous influence of Freud's ideas on twentieth century thought.

1.2.2 Learning Theory

Learning theory or Behaviourism emerged early in this century as a reaction to the methodology of introspection. As John Watson stated in his declaration of Behaviourism in 1913, the goal of psychology should be to predict and control overt behaviour, not to describe and explain conscious states. Learning theorists asked questions that could be answered and provided a powerful methodology for answering those questions. Central questions were: can learning take place without explicit reinforcement? Why is a learned response more persistent if it has been only partially reinforced?

Learning was considered all-powerful (given the right reinforcement schedule), and between-species differences were played down. Development, was regarded merely as the steady accumulation of stimulus-response associations and the methodologies were transferred from rats and pigeons to children with almost no alterations, including operant and classical conditioning and even maze learning (Hicks and Carr, 1912). A review of children's learning (Munn 1954), concluded that the laws of learning are the same in children as in other populations and that children learn faster than rats but slower than college students. These sorts of results led Skinner, in his novel, Walden Two(1948), to propose that children in his utopian society would be raised by behavioural engineers: specialists in operant conditioning!

Children soon showed that they were unlike rats in many ways, forcing the learning theorists to recast them as "rats with language"! Children could label attributes of objects such as their size and colour, and use these labels to guide their learning. Learning seemed to be under cognitive control in other ways as well. Attending to relevant information, forming hypotheses, and generating strategies increased their speed of learning. The experiments of the learning theorists be-

came more and more abstract and artificial in the doomed attempt to avoid these "contaminations" of the "pure" learning processes themselves.

Discontent set in after hundreds of these studies had still not led to a satisfactory account of memory and learning, and researchers began wondering whether the "contaminations" might be more interesting than the processes they contaminate. Also, new evidence highlighted the importance of biological predispositions; e.g. rats can associate nausea with a taste but not with a light or a sound (Garcia and Koelling, 1966). Furthermore, Chomsky(1959) showed that Skinner's account of language was severely lacking, because learning approaches could not possibly explain the acquisition of a skill as complex as language.

1.2.3 Ethological Theories

Modern Ethology began in the 1930s with Konrad Lorenz and Niko Tinbergan as a reaction by young zoologists to studying dead animals. In contrast to Learning Theory, Ethology sees animals as active organisms living within a particular ecological niche, not as passive organisms prodded by stimuli. Ecology also stresses the importance of species-specific innate behaviour – "an animal lacking "instincts" altogether would be a complete slave of conditioning", (Guiselin and Scudo, 1986). Innate behaviours are considered similar to organs of the body, in that both are essentially the same in all members of the species, are inherited and are adaptive (Lorenz, 1937).

Ethology does consider learning to be important but does not appeal to universal all-powerful laws but to species-specific learning predispositions. To a great extent the brain is "like an exposed negative, waiting to be dipped into developers fluid", (Wilson,1975). Species differ in which aspects of their behaviour are modifiable, in what kinds of learning occur most easily, and in the mechanisms of learning. Learning predispositions also include sensitive or critical periods during which an animal is biologically ready to acquire a new behaviour.

The importance of developmental research is thus seen as follows: "the developmental point of view is basic to an understanding of how evolutionary and ecological parameters are achieved in individuals and groups", (Gottlieb,1979).

The most important figure to bring ethology to the attention of developmental psychologists was John Bowlby – a reformed Freudian. From work on the large numbers of orphaned children after the second world war he proposed that early social "attachment" between infant and caretaker is crucial for normal development. This replaced the Victorian attitude that "children should be seen and not heard", and is now totally accepted. Bowlby argued that this need has evolved because it promotes the survival of helpless infants by protecting them from predators or exposure to the elements, (Bowlby, 1969). Most important to human infants are the signalling mechanisms such as crying, babbling and smiling. These behaviours communicate the infants needs to the mother.

The major criticism of the ethological approach is that it describes more than it explains. The mechanisms of development are entirely maturational even to the extent of determining the critical periods. Concluding that a child acquires a certain behaviour because it is in a critical period, does not explain why the organism is pretuned to certain experiences at one time rather than another. Furthermore, it is impossible to test the claims of ethological theory, either because the critical experiments are unethical, such as depriving a child of certain stimulation during the critical period, or because the experiments would have to run over an evolutionary timescale.

1.2.4 Vygotsky and the Contextualists

Rather than focus on the child, Contextualists view a child-in-context participating in some event as the smallest meaningful unit of study. The mind is not a constant universal organism operating in a vacuum but is inherently social. Context affects the child and the child affects the context. Because of this interrelatedness, looking at a child whilst ignoring the context distorts our concept of the nature of children. The fusion of children and contexts may not seem like a new concept,

because other theories emphasize the interaction between children and their environments. However, in these other theories the child and the context are seen as separate entities which interact, whereas by contrast, the Contextualists view this separation as artificial and distorting.

Perhaps the most important idea that Vygotsky contributed to Developmental Psychology was the concept of **zone of proximal development**. Vygotsky defined this to be the distance between a child's "actual developmental level as determined by independent problem solving" and the higher level of "potential development as determined through problem solving under adult guidance or in collaboration with more capable peers", (Vygotsky, 1978). The more competent person helps the child by means of prompts, clues, modelling, explanation, leading questions, discussion, joint participation, encouragement, control of the child's attention and so on. As Vygotsky explained, "learning awakens a variety of internal developmental processes that are able to operate only when the child is interacting with people in his environment and in cooperation with his peers", (Vygotsky, 1978).

The cognitive result of this interaction with adults and peers is an internalisation of the interaction, thus, thinking is always social and reflects the culture in which the dyad operates. As Vygotsky expresses it, any intellectual function acquired during development "appears twice, or on two planes .... It appears first between people as an intermental category, and then within the child as an intramental category. This is equally true with regard to voluntary attention, logical memory, the formation of concepts, and the development of will."

The contextualists theory obviously relies heavily on language as no lesser type of communication could possibly account for the powerful effects on high-level mental processes. In fact, Vygotsky believed that thought and language were inextricably linked by the age of two. But the contextualist story has not successfully addressed non-linguistic development, nor the mechanisms that are required for responding to prompts, joint attention, learning from observation, and collaborative dialogue. It is these that define the zone of proximal development and until we understand them it will remain a vague concept.

As Brofenbrenner(1986) put it: "in place of too much research of development *out of context*, we now have a surfeit of studies on *context without development*". We must understand development from the bottom up given that it is an incremental process. At the moment, Vygotsky's theory is rather weak in that it explains higher mental processes in terms of even vaguer concepts such as culture and social context. It will remain weak until it can be grounded by the fundamental developmental mechanisms that allow these influences to occur in the first place.

1.3 Cognitive Development is a Creative Adaptive Process

Piaget's theory of development can be seen as the middle course between Behaviourist and Ethological theories. We are neither slaves to our environment nor slaves to our innate instincts or programs. Although Piaget believed that development cannot be merely the gradual unfolding of innate structures by maturation, triggered or released by experience, he did accept, following Kant(1787), that we need knowledge structures to make sense of the world. Unlike Kant, however, Piaget rejected the idea that these knowledge structures must all be innate. Piaget believed that the process of development was an interactive creative process whereby the developing system and the environment it inhabits interact in a dynamic way, such that the system is moulded by the environment and moulded for that specific environment.

Piaget did not come to these conclusions without reasons. Trained as a biologist and with a deep interest in philosophy, his aim was to reconcile epistemological problems with biological constraints. The biologist might ask: how does knowledge contribute to the adaptation of the organism, and how has it evolved throughout the phylogenetic scale? And the epistemologist wants to know: how is knowledge possible and what types of knowledge are essential to our view of reality? His reason for studying cognitive development was the biological observation that ontogenesis often mirrors phylogenesis, which explains his statement: "psychology

is the embryology of intelligence". The questions of knowledge from the cognitive developmental point of view become: what types of knowledge do humans have and how do they develop? Piaget considered that the answers to the questions in one field must be broadly the same for the others. Thus he saw parallels in the schisms in each of the three different subjects: Darwinism vs Lamarckism, Empiricism vs Rationalism, and Behaviourism vs Gestaltism. In each the former posits genesis without structure, and the latter posits structure without genesis. Piaget's answer in each case was *structure with genesis.* From this comes the basic idea that the child constructs its own reality; the child does not only *discover* structure in the world, nor does a child merely *impose* structure onto the world. Knowledge is a useful thing that allows us to to exploit our environment. It does not have to be a 'true' reflection of the 'real' world out there, but precisely because it must be useful it must be applicable to the real world, which explains why Piaget considered action to be so important. In fact, Piaget stressed that only through an agent's actions, whether overt or internalised, is knowledge possible.

Once we have accepted this view of development as the self-organisation of an autonomous agent then the questions to be answered become clear.

- What abilities do we have to begin with?
- Do specific abilities come in a specific order and if so which order?
- Does the process ever end and if so when and where?
- What drives the process?

It is important to note that the ontological and ontogenetical questions are not posed merely to determine a timetable for cognitive development but to gain pointers to the underlying mechanisms and forces that control the process.

Piaget decomposed the interactive process of cognitive development into two parts: **assimilation** and **accommodation**. Assimilation describes how the developing system takes in the world and fits it into existing knowledge structures. Accommodation describes how the existing knowledge structures are changed by

the world. Although Piaget split the interactive process into these component parts for the sake of description he believed that they always occurred together and are usually inseparable. He dubbed the process **equilibration**. This process is creative. Piaget viewed it like a Hegelian dialectical process whereby the knowledge structures pass from thesis to antithesis and to synthesis, leading to a spiralling generation of new knowledge structures. Such a process requires something to drive it. The analogy to a dialectical argument shows clearly that for Piaget, the driving force is truth preservation. This naturally leads to the view that knowledge is based on logic because truth preservation is what logics are good at. Practically everything else in Piaget's theory of development follows from these ideas, but as will be shown in what follows, this leads to insurmountable problems. That does not mean, however, that we must reject the idea of development as a creative adaptive process.

1.4 Piaget's Theory of Development by Stages

Piaget's wrote a huge amount but his prose is not the easiest to decipher. For good introductions to Piaget's theory see Boden(1979) and Donaldson(1978, Appendix A).

Piaget considered development to proceed in stages. He considered there to be three main periods each characterised by groups of logical operations. Each stage built on the previous one and each successive stage gave the child more control over its environment. The first stage, called the sensori-motor period lasts from birth to about eighteen months. Piaget believed that new born infants have a very limited range of abilities, i.e. they have few innate structures. These reflex responses, such as sucking, swallowing and the like, are not isolated responses but are embedded in a wider range of spontaneous rhythmic activities. The infant also has the capacity to set in motion the complex processes of accommodation and assimilation which transform the rigid reflexes into surprisingly flexible patterns of behaviour before the first period is over. The most important transformation

that occurs during this period but which cannot be directly observed is that the infant develops the ability to distinguish between themself and the world. By the end of this period the child has constructed the notion of a world and objects, which are independent of them and their actions.

The second period, known as the concrete-operational period, lasts from eighteen months to eleven years. Piaget considered this period to be split into two substages. The first of these, known as the preoperational period, lasts till seven years old during which the 'concrete operations' are being prepared for. During the second period the operations are established and consolidated. Tasks such as conservation and class-inclusion are designed specifically to elicit responses that allow us to tell whether a certain operation is functioning or not; e.g. when a child reasons that the number of a set of objects remains the same although its arrangement in space has been altered, that child is said to do this by understanding that the original arrangement could be reached again merely by reversing the movements that changed it. Thus the the child's thoughts are reversible.

The preparatory work that has to take place during the preoperational period before the operations come into being consists mainly in the child's capacity to represent things to himself. Group structure already exists by the end of the sensori-motor period, but on a practical level only. The next step is to internalise it. Internalisation is not just somehow taking it as a whole, but means rebuilding on a new plane. The work of the sensori-motor period has to be done all over again, but now the building blocks are symbols in the mind. A child of two or three can arrange objects in a row, space them out, and put them together again; a child of seven can think of doing these things.

The third period is known as the formal operational period. The thinking of this period, once it has been consolidated is that of the thinking adult. Its most marked feature is the the ability to think logically starting from premises and drawing conclusions that necessarily follow. This leads to the ability to plan systematic experiments in which a child will realise, for example, the value of holding one thing constant, while letting others vary, and thus formulate general rules based on experimental findings.

The main difference between the formal operational period and the concrete operational period is that the child develops from being able to manipulate things in the mind to being able to manipulate ideas or propositions in the mind. This constitutes a shift in the relation between what is real and what is possible; the formal operational child can explore all the possibilities.

1.4.1 Piaget's Theory of the Development of Transitive Inference

Transitive inference was first brought to the attention of psychologists by Piaget. It was particularly important within his theory of cognitive development. For Piaget, intelligence is built on logic and he argued that being able to make transitive inferences with abstract symbols marked the end point of intellectual development. Thus a child who can solve the problem:

Edith is fairer than Susan. Edith is darker than Lily. Who is the darkest?

has fully developed to the level of an adult. Piaget claimed that the end-point of logical development usually comes at about the age of 11 years old.

Piaget (Piaget & Inhelder, 1973) analysed the development of transitive inferences in terms of the childs progressive understanding of relations. According to Piaget, the child progresses through a number of substages from a categorical to a relativistic conception of relations. Operating according to a categorical conception, children at the preoperational stage tend to regard relations as absolute properties or attributes of things. For instance, "larger" and "smaller" are taken to be two mutually exclusive attributes of objects. Therefore, the child cannot understand that an object can be both larger than one object and smaller than another.

One of the essential characteristics of concrete operational thought, according to Piaget is its reversibility. In the area of relational thought, reversibility is the ability to coordinate inverse relations. This reversibility may take two forms:

1. the transformation of a relation ($A > B$) into its inverse ($B < A$) and back.

2. the coordination of inverse relations that are not inverses of one another around the same term - for instance, the understanding that an object B can be larger than A and smaller than C.

This second form of reversibility is particularly crucial to an understanding of transitive inference, as it underlies the relativistic conception of relations, which is essential for the coordination of premise relations around their common term.

To summarize, the ability to make transitive inferences relies on the understanding of relations and the ability to manipulate these relations in the head. Piaget claimed that this cannot be done by children under seven years old when it is concrete objects that must be manipulated, such as rods of varying length, and cannot be done by children under eleven years old when it is propositions that must be manipulated, such as the Edith, Susan, Lily example.

1.4.2 Is Piaget's Theory Teleological?

Piaget's theory of rationality might be paraphrased as follows: we need truth preservation procedures to allow cognitive structures to be built so we can make sense of the world in order to obtain truths which ground our reasons for doing what we do! It seems as if truth preservation is both the driving force and the end-point which would make the theory teleological. This however is unfair. Piaget was not committed to logic, but chose it as the most powerful system around at the time. He may have considered truth preservation to be the only possible driving force, but logic need not have been the end-point. He certainly did not believe that his theory should be constrained by logic.

1.5 Why Should Cognitive Development Proceed in Stages?

The developmental tasks devised by Piaget give a snapshot of a particular competence at a particular time, so they cannot shed any light on whether the underlying process is continuous or not. So it was not empirical evidence that led Piaget to believe that development was discontinuous. Nor did he believe that the stages were just an artifact of a continuous process that happens to be extremely slow at some times and faster at others. The importance of the stages lies in the fact that Piaget believed that intelligence relied on logic. For the purposes of reasoning, however, half a logic is not much use! So each stage has a complete logic and the higher the stage the more powerful the logic.

1.5.1 The Transition Problem

The various tasks that mark transition from one stage of development to the next give some of the most robust results in the history of psychology, but Piaget was extremely vague about the mechanism of the transitions. Piaget was adamant that the change was not due to maturation; maturation can do no more than temporarily limit the possibilities. He argued that developing from one stage to the next was a matter of restructuring on a new plane what had been achieved at the previous level. He used the term **reflection** for this transition to convey both that it is some sort of mathematical transformation and that it gives rise to increased self-awareness. But Piaget never found a useful formalisation of terms such as "refection" or "equilibration"; thus they remain as no more than metaphors.

Nobody has been successful in showing how a less powerful logic describing and embodying reasoning at a lower stage, can give rise to a more powerful logic at a higher stage. The best attempt to specify these magical transitions, involved formulating explicit transition rules. These transition rules, however, turn out to be exceedingly powerful in their own right; perhaps more powerful than either

of the logics involved in the transition. Furthermore, these meta-developmental rules cannot have developed themselves so where could they have come from? Furthermore, if we have these highly sophisticated rules from birth then why do we not apply them immediately and become fully developed at once? Finally, the very idea of transition rules goes totally against the spirit of Piaget's theory. Piaget considered development as an open and creative process which cannot be reconciled with innate transition rules.

1.6 Reviewing Piaget's Ontology

If rationality is based on logic and it develops over a long period of time then it is seems natural that Piaget's hierarchy of logical manipulations and structures develops in the way that he suggested, and that transitive inference is the end-point of cognitive development. Given that the transition problem causes insurmountable problems for the view that logics develop, Bryant and Trabasso(1971) challenged the ontological status for transitive inference that Piaget had claimed.

Bryant and Trabasso argued that Piaget's test for the concrete operation of transitive inference was methodologically flawed and once the flaws had been removed children as young as four could make transitive inferences perfectly adequately. In the task for testing the concrete operation for transitive inference, children were presented with two pairs of coloured sticks. In the first pair, for instance, a red stick would be longer than a blue stick, and in the second pair the blue stick would be longer than a yellow stick. The children would then be asked which was longer, the red or the yellow stick. There are two major criticisms of this paradigm. Firstly, the children may be failing on the task not because they could not perform transitive inference, but because they could not remember one or both of the premises (red is bigger than blue and blue is bigger than yellow). Secondly, the children may be succeeding on the task by using a non-logical labelling strategy instead of transitive inference. For instance, if the child labelled the red stick as big and the blue one as small in the first pair, and the blue stick

as big and the yellow one as small in the second, then the test question involves only the two end-points which are not labelled in a contradictory manner. This allows the children to just read off the answer from the labels. Smedland (1966) had pointed out both of these criticisms, but Bryant and Trabasso (1971) were the first to consider both together in a new version of the experimental paradigm. They controlled for memory by training the subjects on the premises, until they had reached a 90% criterion of success. And they controlled for labelling strategies by increasing the number of items in the series upto five so that there is a comparison that requires an inference but does not involve end-points. For example, in the series $A > B > C > D > E$ the terms B, C and D are all ambiguously labelled as being bigger than one thing and smaller than another, which allows a comparison between B and D which could not be answered significantly better than chance using the labelling strategy.

Another factor that Trabasso attempted to control for was language comprehension. He did this by presenting each premise in terms of both comparatives. Trabasso and Riley (1974), found that whereas 87% of four year olds reached the criterion (for memory of the premises) when both comparatives were used, only 35% reached it when only one was used. They explained this in terms of the type of premise encoding that each condition encourages. When the child is only given the premises in terms of "longer", they tend to use a nominal encoding (reflecting categorical conception), i.e. they label each stick, but because four sticks are labelled as "long" it is difficult to distinguish between the premises. By giving the children both comparatives, Trabasso and Riley assumed that they were eliciting the childs relativistic conception of the relation. The reason that the bicomparative elicits the relativistic conception of the child, they argued, is because it does not allow a nominal encoding in which there are no contradictory labels.

Given these alterations to the paradigm the following results were obtained: the percentage of transitive responses on the BD comparison were 78%, 88%, and 92% for four, five and six year olds respectively. This is strong evidence that transitive inference is much closer to the lower ontological bounds of the system, (in the formal systems sense described earlier). It is difficult to judge the performance of

children younger than four as there are so many factors that might lead to their failure which do not involve the inability to make transitive inferences. Therefore, with this paradigm it impossible to tell just how primitive the ability is.

1.6.1 Is Transitive Inference Based on Logic?

If transitive inference, (and by implication, logic), is innate, then what role does development play? Very little is seems, but the very paradigm that was used to make this claim also provides evidence that transitive inference does not rely on logic at all. If transitive inferences were made by logic then it would be expected the the greater the number of inferences required, the longer it would take to make the choice. However, in fact the reverse turns out to be the case. The greater the distance between two items the faster the subjects respond. This is known as the **symbolic distance effect**.

Controlling for language by increasing the linguistic component of the task, (having two comparatives instead of one), seems strange to say the least.

McGonigle and Chalmers, (McGonigle and Chalmers, 1977; Chalmers, 1977; Chalmers and McGonigle, 1984; McGonigle and Chalmers, 1986; McGonigle and Chalmers, 1992) further developed the paradigm to make it truly non-linguistic. To make absolutely certain that no linguistic component was present, the experiment used monkeys as well as children as subjects. Instead of using coloured rods of different length they showed subjects pairs of coloured tins, one of which was heavier than the other. Both the monkeys and children had no problem learning the premises and both showed a strong transitive bias on the remote pairs. There was also a clear symbolic distance effect.

Ferson, Wynne, Delius and Staddon(1991), trained pigeons on a 5-term transitive inference task. The pigeons could not learn the premises to the same precision as children or monkeys, and yet they still showed a strong positive generalisation to the remote pairs.

The fact that monkeys "not well known for their logical skills", show the same level of performance as six-year old children, and pigeons, (hardly the brainest of

animals), also show a strong transitive bias, is the most conclusive evidence yet that the underlying mechanisms are ontologically primitive and do not require any sort of logical manipulation in the head.

Once the theory that transitive inferences are based on logic is rejected, the empirical paradigms must change in order to see what the skill is based on. So instead of just teaching the premises and testing the inferences, new transfer tests were added. Having taught the monkeys a five term series, McGonigle and Chalmers(1986), then gave subjects a post-test of sets of three coloured tins (triads). One might expect that the *bcd* triad would produce more responses to *b* than the *bd* pair, because if subjects are coordinating the information in the premises then it should be easier if the central item around which the coordination must take place is present. However, on all the triads they found that subjects gave a distribution of responses; for example, for the *bcd* triad subjects typically give all three possible responses roughly in the ratio 50:33:17. So there is still a transitive bias but this bias is reduced as the number of items increases.

1.7 What is the Basis of Rationality?

1.7.1 Logic

A logic is just an artificial language in which certain statements can be expressed, and in which other statements follow necessarily from the Laws of Entailment that make up part of the definition of the logic. There are an infinite number of possible logics, and if we are to accept the Church-Turing thesis, there is a logic that can describe any system that is describable. We might conclude from this that rationality is based on some logic but we just do not know which one yet. This is absolutely wrong. I would hope that we could describe the basis of rationality in plain english but I would never make the claim that rationality is based on english.

It also makes little sense to investigate rational processes by exploring the properties of different logics. This is true precisely because there are an infinite number of logics. Thus, logic cannot constrain rationality but it does constrain research on rationality. Some logics are easier to develop and understand than others and so will be likely to be explored sooner, but this is exactly the wrong sort of constraint. Research should be constrained by the thing being investigated, not by some representational medium of description.

1.7.2 Mental Models

Johnson-Laird (e.g. Johnson-Laird, 1983), claimed that much of our reasoning was not based on logic. He proposed that people build models 'in their heads' which represent a possible state of affairs, and produce answers to problems by simply reading them off from these mental representations. The mental models that people can build are highly restricted and these restrictions are the basis of two important differences between Mental Model Theory and more traditional Model Theoretic accounts. Firstly, the 'things' that are represented in the models are not abstract symbols but representations of specific objects. Johnson-Laird claims that this can explain why our reasoning is so often affected by the content of what we are reasoning about, (e.g. Wason and Shapiro, 1971), but he never actually explains adequately how. Secondly, and more importantly, Mental Model Theory attempts to account not only for successful reasoning but also for unsuccessful reasoning. Many problems require people to build more than one mental model in order for them to come up with the right answer but people find it hard to build multiple models, and so they often build fewer than they need and read off the wrong answer. So whereas the traditional logical accounts were theories of competence, Johnson-Laird's Mental Model Theory attempts to account for both competence and performance.

Turning to transitive inference, Ehrlich and Johnson-Laird(1982), predicted that if the premises are indeterminate, $(a > b, a > c, b > d, c > d)$, then subjects would have to build multiple models and therefore they would find the task diffi-

cult if not impossible. Wright(1985), gave subjects a large number of transitivity problems, some of which were determinate and some of which were not. He found the opposite to that which Ehrlich and Johnson-Laird had predicted; namely, that subjects had no problems with indeterminate relations at all. They coped just as well with indeterminate relations as with determinate ones, and often, they did not even notice the indeterminacy.

More recently, Johnson-Laird, (Johnson-Laird and Byrne, 1993), seems to have adopted the position that when subjects find indeterminacy easy it is because they build just one model, and when they find it hard it is because they are trying to build more. But a scientific theory is supposed to predict these things not just be able to accommodate any eventuality!

1.7.3 Transitivity of Preferences

Economic theory requires that agents in the market place be rational, (Edwards, 1954). Part of the economic definition of rationality is that preferences should be transitive. The cost of not being transitive is that an agent can become a 'money pump'. Consider an agent that has a circular set of preferences, i.e. it prefers object a to object b, object b to object c, and object c to object a. According to economic theory, a rational agent will be prepared to pay money to exchange an object for a more preferred one. Therefore, an agent with the above preferences and who already owns object a will pay money to exchange it for object c, more money to exchange c for b, and yet more money to get back a again, and so on *ad infinitum*. This is obviously disastrous for the agent and it can only be avoided by imposing a minimum of weak stochastic transitivity over the premises.

Thus, Tversky(1969) described transitive inference, not as the end-point of the development of logic, but as "the simplest and probably the most basic principle of choice".

Unfortunately, the story is a bit more complicated than this. It is simple to construct situations that not only make subjects produce intransitive preferences but also makes it seem reasonable to do so (Tversky, 1969). For example, imagine

you want to buy a computer. Suppose there are two criteria that are important to you: the computers' speed and the amount of available memory. Suppose also that speed is the more important criterion and that you will only choose on the basis of memory if the difference in speed between two computers is less than 2MHz. Now consider the three computers shown in the table below.

	Speed (MHz)	Memory (K)
Computer A	6	256
Computer B	5	512
Computer C	4	640

The difference in speed between A and B is only 1, so you would prefer B on account of the greater memory. Similarly, you should prefer C to B. However, because the difference in speed between A and C is not less than 2, you should prefer A to C.

The above example requires binary comparisons and more than one criterion. This scenario may seem artificial, but surely these sort of complex choice situations are ubiquitous in our everyday lives. Economic theories presume that every choice can be evaluated in the same currency, but choices such as "Do I try and finish this thesis as soon as possible or do I go and have a nice curry?" are difficult precisely because hunger and the desire to finish working cannot be measured in the same currency.

Finally, if imposing transitivity on any set of preferences is the definition of rationality, then we are forced to conclude that often people are irrational. But the tests given to people which show intransitivity do not show that peoples' choices are unprincipled. If our theory of rationality becomes normative then it ceases to be of any psychological interest. The fact that people's choices can be illogical and intransitive, yet principled is of psychological interest, precisely because the principles concerned are suggestive of the underlying basis of rationality.

1.7.4 Does Rationality have a Unitary Basis?

The only useful definition for rationality at the present time must be: rationality is the answer to the question "Why do people do what they do?" Accepting that we do not know the answer is a huge advance, because now we are free to explore all the possibilities. The first thing to consider is whether rationality has a unitary basis at all. There is no doubt that I use logic sometimes in order to draw certain conclusions, and I am quite willing to believe that in solving syllogisms I use something akin to mental models. Undoubtedly I spend much time evaluating different options open to me. All of these abilities to solve different sorts of problem need not all rely on the same underlying mechanism, and their ontological status is unclear. For example, it is not surprising that I often use logic because I am the product of a western educational system in which logical argument is central.

Considering that our fundamental rational mechanisms are the product of evolution, it would be very surprising indeed if the basis of rationality turned out to be a huge monolithic system; evolution selects for specific solutions to specific problems. Which raises the question of how such a general purpose organism such as *homo sapiens* evolved. The idea that we are general purpose organisms is highly anthropocentric to begin with; there are many things which we are very bad at, like flying. However, it must be conceded that we are highly adaptable and can solve a very large number of different types of problem with which the world confronts us. There are two ways of being a general problem solving agent. One is to have a general problem solving ability, and the other is to have lots of specific problem solving abilities and the capability to reduce any problem to one we already know how to solve. This specific generality is the sort that might be selected for, (e.g. Sherry and Schacter, 1987;Brooks, 1991).

1.8 Re-Starting from the Beginning

The problem of cognitive development is fundamentally the problem of systems growth: what advantages does it convey, how is it done, what problems does it bring with it and how do we cope with them? As stated at the beginning, the purpose of development must be to allow the agent to better adapt to its environment. Therefore, this process must have input from both the environment and the agent. Hence, the process is interactive. Given this interactive process, we then ask: what drives it and controls it? To answer this question it helps to know exactly what it is that develops. Piaget thought it was logic that develops , that transitive inference relies on logic and so marks the end-point of development, and that the driving force is truth preservation. We now know that transitive inference is close to the lower ontological bounds of the system, that it does not rely on logic, and that truth preservation as the driving force is highly problematic. Which leaves us to determine a new ontology and ontogeny, and to think again about the underlying mechanisms and driving forces of development.

There is a great deal of research to be done, but a good place to start is to re-examine transitive inference. To begin with, why is it so primitive?

1.8.1 Imposing Transitivity as a Default Assumption is a Pragmatic Thing to do

Any quantity that can be measured on a monotonic scale has a transitive relation associated with it, and as quantities such as height, length, weight, time, distance, and money, (to name but a few), figure prominently in many of the decisions we must make it is not surprising that this inferential ability is fairly primitive. This idea that pragmaticism governs the ontology of the ability to make transitive inferences might be problematic if transitive inferences were particularly difficult, but they are not. What seems much harder than making a transitive inference is

knowing whether the underlying relation is transitive or not, and so whether the inference is appropriate or not.

Whereas it does not make sense to claim that we must impose transitivity over our set of preferences in order to be rational, it makes a great deal of sense to impose transitivity even when we do not know that the relation we are reasoning about is transitive. We need some basis on which to make decisions and if we do not have all the available information, then making the assumption that the underlying relation is transitive is probably the best assumption we can make. In other words, transitivity of preferences is not the basis of rationality but making transitive choices is a rational thing to do.

1.8.2 How Does this Fit in with our Ideas of Development?

One could argue that the ordering of choices is nothing less than a construction of a workable version of reality. Without an ordering of choices we have no basis for making a decision, so our version of reality is unworkable. And to achieve the best ordering, the process must be informed by the world.

Of course, not all relations are transitive, and therefore a simple ranking of options is not always the appropriate thing to do. So can we cope with intransitive relations and if so how? Assuming we can, then the Piagetian solution is that through interaction with the environment we construct the appropriate means for dealing with whatever relation we are attempting to reason about. If the relation is transitive then we may construct a simple ordering of the items, and if it is not we construct something else. Of course, many relationships in the world may be indeterminate, in the sense that we do not have sufficient information to know whether they are transitive or not, and then it makes sense to assume transitivity to begin with, but this assumption should be defeasible in the case that we obtain further information that the relation is not transitive afterall. Assuming transitivity as a default yet defeasible assumption, allows an agent to exploit its environment as much as possible whilst still remaining adaptive.

To explore these possibilities we must first ask whether there is any evidence of subjects self-regulating in transitivity tasks, and then see how subjects cope with intransitive relations.

1.8.3 Evidence for Internal Regulators

McGonigle and Chalmers(1992), retrained monkeys on a previously learnt 5-term series, and reproduced the drop in performance to the triads. The subjects were then given the triads in a dense training phase, but still without discriminatory feedback. Most of the subjects self-regulated their performance, such that they always chose the highest colour in the series. Those that did not self-regulate, were then given discriminatory feedback. All of them learnt to choose the highest colour in the series in a very few trials. This is the first example of animals learning in the absence of discriminatory feedback since the latent learning experiments of the 1950s. It strongly suggests that monkeys have their own internal regulators.

1.9 Can we Cope with Circular Relations?

Every researcher in this field mentioned so far assumed that all subjects understand *a priori* that relations such as length and weight are transitive relations. But many of the experimental paradigms actually obscure the scalar properties of these relations by avoiding giving absolute quantities, and by making the relative quantities not directly observable. One of the problems facing the subjects in these experiments therefore must be to work out whether the relation is transitive or not. Consider the three term transitivity task where subjects are taught to choose a red rod over a blue rod, and a blue rod over a yellow rod. Although the lengths of the rods are obscured, the subject must realise that somehow length is relevant, and furthermore that length is a scalar quantity and therefore comparisons of length are transitive. All this is totally independent of whether the subject knows how to make transitive inferences or not. The subject does not know that they are not about to be taught to take the yellow rod over the red rod. Relations such as *to-*

the-left-of would appear to be transitive relations and yet, as Johnson-Laird(1983) has pointed out, if you imagine King Arthur's round table, everyone is both to the left and to the right of everyone else. Any system for learning relations must be robust to this sort of scenario; i.e. it must be able to produce some reasonable response. The obvious response is to consider how far round the table in each direction it is necessary to go. Thus at a table with five people there will be two people to the left of everyone else, (either one or two places), and two people to the right.

1.9.1 Empirical Evidence on the Ability to Cope with Circular Relations

Gillan(1981) trained chimpanzees on a six term series. They managed to learn the premises and they generalised correctly to the remote pairs. Gillan then further trained one of the subjects on a circular relationship made up of the original six term series but adding a sixth premise in which the subject was trained to choose the stimulus which had never been rewarded, over the stimulus that had always been rewarded. Gillan concluded from the results that chimpanzees could be trained to respond correctly on a circular series but that they failed to generalise appropriately once the end-points had been taken away. There are three major flaws with the experiment that make it impossible to accept Gillan's conclusions without further research.

Firstly, only one subject was trained and tested on the circular relation, hardly a huge sample from which to determine the limits of cognition! Secondly, the subject was not taught a circular relation from the beginning but taught a serial relation that then had its ends closed. Therefore, one could only conclude that a chimpanzee cannot transfer its knowledge of a serial relation to a circular relation that is made up of the same components. Thirdly, the data are not particularly convincing to conclude anything. Gillan based his conclusions on the results in table 1–1.

Adjacent Pairs						Nonadjacent Pairs		
A-B+	B-C+	C-D+	D-E+	E-F+	F-A+	BD	BE	CE
7/8	5/8	7/8	8/8	8/8	8/8	3/12	8/12	7/12

Table 1–1: Results from the subject on a seven term closed relation, (Gillan,1981)

The subject only scored five out of eight on one of the crucial premises. This is not even significantly better than chance so it is not reasonable to conclude that the chimpanzee could learn a circular set of premises let alone make any conclusions based on the transfer of knowledge which it does not necessarily possess.

The subject was taught the six term series (made up of five premises) in a total of thirty three sessions. Only fifteen sessions were given on the circular relation. The circular relation has six premises and although only one of them must be learned from scratch, as Gillan points out himself, all the premises must be integrated together somehow. It seems highly realistic to assume that integrating a circular set of premises would be significantly harder than a linear set. This is especially true if five of the six premises have already been integrated together in an inappropriate fashion to integrate the sixth. We are left with the possibility that had the subject been trained further on the circular relation, then perhaps they would have learned the premises to a suitable criterion which would have made the transfer tests interpretable. This sort of *what-if* question is one of the main advantages of computer modelling, where the cost of running the model even for another hundred sessions is minimal.

Fortunately, there is data which is slightly more interpretable. Ferson, Wynne, Delius and Staddon(1991), working with pigeons, also made the mistake of trying to make conclusions about their pigeons ability to learn circular relationships, by first training on a serial relation and then adding the final premise to make it circular later. Their results, however, are clearer.

Of the four subjects which started, one dropped out, one could not learn the premises better than chance, but two did manage to learn the circular set of premises. The two successful pigeons did not generalise in any systematic way. It is difficult to draw any conclusions about the pigeon that dropped out, and we

must not read to much into the failure of one of the pigeons to learn the circular set, because again the number of training sessions given on the circular relation was significantly lower than the number previously given on the serial relation. Despite this, two pigeons did learn the circular set of premises significantly better than chance. Ferson et al, do not like this result because it happens to contradict their particular model of transitive inference. They explain it away as rote learning! But there is a much more interesting explanation.

Transitive Inference as a Defeasible Bias

The pigeons are not learning the premises of a transitive relation, but integrating information gained into a knowledge structure that allows generalisation to novel situations. The premises of a serial relation lead to a knowledge structure in which transitivity is an automatic generalisation. When the relation is changed from a serial to a circular relation the knowledge structure changes, so the transitive bias is defeasible. Of course, 'knowledge structure' is used here to refer to specific task-solving strategies and not some group or grouping of abstract relations as Piaget used the term, but the characterisation is essentially Piagetian, even if all the details have changed.

1.10 Subjects Performance Interpreted as Micro-development

As the subject gains information from feedback for specific choices, they try and integrate this information with all the rest of the information they already have. The new information causes the knowledge structure to change slightly. In other words the knowledge structure *accomodates* to the new information. The way the new information is integrated depends on the existing knowledge structure. In other words the the knowledge structure *assimilates* the new information. Another example of assimilation is the transfer to the remote pairs. The subjects have not seen these novel pairs before, but because they can assimilate them into the

existing knowledge structure they can generalise. When the subjects are taught the additional premise to make the relation circular, they cannot assimilate this information into the existing knowledge structure, but eventually the knowledge structure accomodates to it, and can respond correctly.

Piaget analysed the development of transitive inference in terms of the child's progressive understanding of relations. This, I believe, is essentially correct but surely Piaget got the scale completely wrong. He was referring to macro-development of relational understanding in general, whereas under this new characterisation, this 'progressive understanding' occurs on a micro scale and concerns specific relationships between specific objects.

1.10.1 Controlling for Memory?

This view of subjects' performance as micro-development, explains some rather strange facts.

Lunzer and Lucas(1977) gave a version of the transitive inference task to children of different ages in which all the information in the premises was perceptually available at the time of testing on the remote pairs. Even nine year olds only showed a very weak transitive bias, which indicates that it is the learning of the premises that is important for young children to make the correct inferences. Thus, it seems as if Bryant and Trabasso did much more than just control against lack of memory. Understanding the process of learning appears to hold a vital key to understanding the ability to make transitive inferences.

Consider the question: why do children and monkeys take so long to learn the premises? To answer it as Trabasso did - that it is because of the deficiencies in working memory - is like attributing the causal effects of morphine to dormative properties; i.e. circular and question-begging. The more Piagetian explanation is that only by a slow and incremental learning procedure can the assimilatory and accommodatory processes interact efficiently to produce the appropriate cognitive structure for the specific task in hand.

We have arrived at a rather ironic state of affairs. Piaget devised the transitive inference task to give a snapshot of logical competence, and as such the task could not give any insights into the process of development. The task has now been modified, (largely motivated by researchers' desire to refute Piaget), such that it requires the subjects to self-regulate their performance whilst interacting with the task. Viewing the task situation in terms of the subject "equilibriating" with the task implies we should model subjects in terms of developing systems. Thus the task has moved from being part of a developmental study to being part of the study of development.

Chapter 2

Mechanisms and Models

Once transitive inference is no longer considered as the end-point of development and reliant on sophisticated symbol-manipulation, questions about the underlying mechanisms become more open. The fact that very young children, monkeys, even pigeons can make these inferences, suggests that the mechanisms must be primitive, and this has led to a plethora of models being developed. This chapter first considers the purpose of modelling and then reviews some of the more major models of transitive inference.

2.1 Modelling

2.1.1 A Model is a Theory

The word "model" means many different things to different people. I follow Newell and Simon's interpretation, (Simon H.A. and Newell A. 1963), where model is just another word for theory; hopefully a more formal, and less indeterminate theory, but a theory nonetheless. Philosophers of science may still argue about the metaphysical status of a theory but I fall unashamedly into the Classical Realist camp who believe that a theory should be a description of the real world which also *explains* the observable facts. It follows that a theory of development must explain causal relationships. As Cummins(1983) put it:

> Transition theories are not explanatory unless the laws appealed to are causal laws .... Subsumption under a generalisation that is not causal merely summarizes our reasons for believing the change would occur – it justifies our expectations perhaps – but it doesn't explain why the change occurs.

This philosophical position makes clear the appropriate relationship between a model, the system being modelled and the empirical data.

The aim of this research is to understand real complex systems in real environments. In order to do this we need tasks to elicit particular behaviours – unless the system behaves we have no data. Given that the system is exceedingly complex, we need to constrain the behaviour as much as possible. This is done through the design of the experimental task. It is absolutely crucial that the task constrains the behaviour in the right way, otherwise it is likely that the behaviour will not inform us usefully about the system, but will merely be an artifact of the specific task that induced the behaviour. This appropriate constraining is often referred to as the **ecological validity** of the task. Without it, generalisation about the system will not work. This raises the question of how to design an ecologically valid task. The only way to do it is to use a current theory of the system as a guide.

Illustrating, using the transitive inference task as an example will make this clearer. If our theory is that rationality is based on logic then the transitive inference task can be considered a good exemplar of logical behaviour and we would expect to be able to generalise substantially from it. Once that theory is rejected through empirical studies, we see that the task as it stands is no longer so useful – it is now seen as ecologically invalid. If a task is not ecologically valid it means that successful subjects are likely to have succeeded using task-specific strategies which will not generalise to other tasks such as the triadic choice extension.

Crucially, we are not interested in performance on the transitive inference task *per se*, but only in as much as we can generalise from performance on this task to the performance of the system, which we can observe using a wider variety of

tasks. This generalisation is predicted by the model or theory and can be tested empirically.

To summarize, a model is a theory: a theory about a system. This theory guides the design of a task, which generates data which can then be used to evaluate the theory. This suggests that the model must come before the task producing the data, which is absolutely right, but the whole process is a loop, (or a spiral if time is included). The data produced by the task may force changes to the theory, which in turn, produces a re-evaluation of the ecological validity of the task, which then either necessitates changes to the task or at least a reinterpretation of the behaviour. As this process unfolds, the theory becomes more and more constrained (and hence more like the traditional view of a model), and the tasks become more and more ecologically valid. The by-product of this loop is a better description of the system and therefore a better understanding of the observable behaviour of that system.

2.1.2 Computer Simulation

Much of the content of this thesis is about how to deal with complexity, (how and why should an agent invest in it?) . The problem of complexity also faces the cognitive modeller. The system being modelled is complex, so the description of the system will need to be complex, but in order to evaluate a model we need to map this description onto behaviour. As the model becomes more and more complex this mapping becomes harder and harder. On solution to this problem is computer simulation. By translating the model into a computer program and then giving the simulation the same task as the subjects, we generate a set of data which we hypothesize comes from the same population as the data obtained from the subjects. Unlike the subjects, however, we can look inside the model and analyse exactly how the data was produced. Thus, analysis of the model's performance is necessary to achieve the second part of the Realists' requirement for a theory – namely that it should explain the observable behaviour.

A simulation also conveys other advantages. The sorts of experiments modelled in this thesis are extremely time consuming to run, as they are longitudinal studies. Furthermore, as the history of experience of a subject is such a crucial variable and the number of simian subjects is limited, a decision to advance the empirical programme in a particular way, most often cannot be changed. With a simulation, however, it is possible to ask interesting **what-if** questions, which can help direct empirical work of the future cheaply and quickly.

To write a simulation which will run on a computer, requires that absolutely everything is specified. This does not mean that we need to have solved every aspect of cognition, but it does mean that basic assumptions about each aspect of cognition must be made explicitly. For example, all the models presented later in this thesis have a representational layer which is presumed to be the output of much sensory computation. People can criticize the models for the representations I have chosen, but that, hopefully, will lead to useful debate. People cannot criticize these models for being vague.

Finally, once a model has been instantiated as a computer simulation, revision of the model becomes much easier. However, this flexibility presents it own problems which are discussed below.

2.1.3 Model Evaluation

There are three major factors in evaluating a model: adequacy, parsimony and generality. Adequacy refers to whether a model actually captures the data. Parsimony uses Occams razor, to compare two different models with equal adequacy. And generality it what it says: how many different types of behaviour can the model capture. However, these abstract measures of performance are not always independent of each other.

If a model is complex and flexible enough, it will be able to capture any particular set of data, by merely setting various parameters to specific values. Such a model is not interesting as it does not constrain the possible behaviours so becomes untestable. Thus, we seem to have a dilemna: we want our theories to be

as general as possible, but also to be as constrained as possible. This dilemna is resolved by insisting that the model is general in the same way as the system being modelled is general. This highlights the main difference between modelling in Psychology and modelling in Artificial Intelligence. In the latter, the more powerful a model is the better.

Attempts have been made to formalise a notion of the parsimony of a model in terms of the number of free variables that are required to reproduce the data. This has its uses but it treats a model as an entity far too independent from the system being modelled and the task to elicit behaviour. Given a specific system and a specific task then the number of free variables is a quantity that should be minimised, but it is possible to have two models, one of which is the right model of the system but due to the fact that the task is the wrong task, captures the data less well than the other model which is not such a good model. An obvious example of this are the Copernican and the Ptolomeic models of the solar system. We now accept the former as the better model but due to the measurements taken, (which embody the task), the latter fitted the data better. Of course, the Copernican model has many fewer free variables, so is more parsimonious. In this case then, parsimony was treated as more important than adequacy.

It is not easy to rank the three factors (adequacy, parsimony, and generality), in order of importance. It often depends on the specific system being modelled and the current state of theories of that system. In the discussion of the models below, I argue that recent models of transitive inference have been so ungeneral that achieving adequacy is too easy, which has led to rather pointless arguments about parsimony in order to discriminate between the different models. The aim of the modelling in the next chapter is to extend the generality which makes adequacy the most crucial factor.

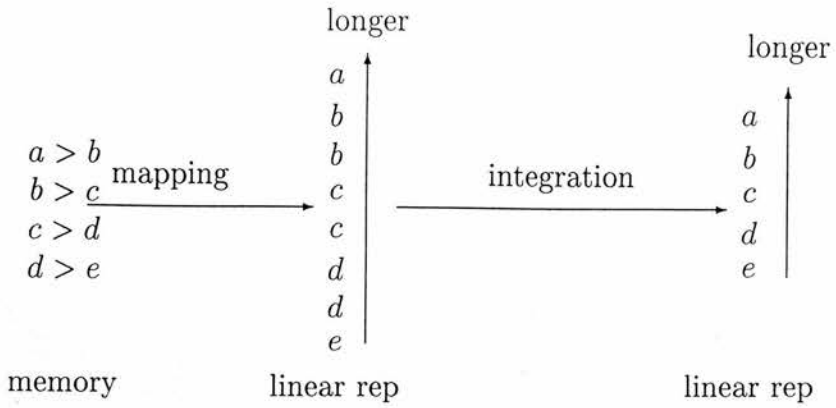


Figure 2–1: Trabasso's Model of How the Linear Array is Constructed

2.2 Internalised Serial Representational Device

Trabasso and Riley(1975), proposed that there was a dedicated representational device for serial relations. During training, pairwise relations are encoded in memory. A representation is then constructed by mapping the pairs onto a mental continuum from the ends inward and then integrating them into individual items. A schematization of this is shown in Figure 2–1.

Notice, from figure 2–1, that the mapping operation is effectively performing transitive inference.

2.2.1 Evidence for an Internalised Serial Representational Device

There were three major pieces of evidence that were used to support this idea. The first was that Trabasso et al found serial position effects whereby premises nearer the end-points of the series were learned faster than those nearer the middle. This effect, so prevalent in list learning, was explained in terms of the end-points being able to be mapped onto the internalised linear array more easily than the items in the middle.

The second piece of evidence was the occurrence of a distance effect: the further apart two items are, the quicker subjects were at responding correctly to the question. This effect had been used by researchers in the "internal psychophysics" movement, (e.g. de Soto, London and Handel, 1965; Huttonlocher, 1968), to argue that subjects used some imagined spatial layout in which the objects to be reasoned about could be positioned.

Further support for a serial representation device came from an experiment from Kallio(1982). Kallio found that serial order cues facilitate the learning of the premises; i.e. if the premises are taught such that the order in which they are presented can be mapped onto the order of size, weight or whatever, then the premises are learnt faster and easier. Obviously, the mapping onto the linear array would be easier if the premises came in the correct order.

2.2.2 Problems with a Serial Representational Device

If there was some internal linear array it is difficult to see how a scanning procedure for making inferences might give rise to a distance effect. In fact, the distance effect suggests a global inspection of the array, and no-one has come up with a non-question begging account of how this might be done.

Why are children, who succeed at the transitive inference task, unable to seriate the items they have just correctly reasoned with, (Chalmers and McGonigle, 1984)? If they have devoted so much effort into mapping them onto a serial order, then surely they must be able to produce this order. Otherwise, the representational device seems of little use outside of this highly specific task.

It also seems strange, to say the least, that subjects should take so long learning the information contained in the premises, when they have a dedicated representational device precisely for that information.

Of course there is no way that a linear representational device can cope with a circular relation. There is, I suppose, the possibility that there is another representational device for circular relations, but how many dedicated representational

devices can we fit in our heads. Furthermore, we then just put the problem one stage back, because we have to decide which representational device to use.

Finally, the monkeys' initial drop in performance on the triads cannot plausibly be accounted for by a linear representational device. Trabasso is vague on the scanning procedure for obtaining information from the linear representation device, but whether this procedure is some global inspection or an item-by-item serial procedure, one would expect that the triadic post-tests would produce a stronger transitive bias not a much weakened one.

2.3 The Sequential-Contiguity Model

Breslow(1981) argued that, whether transitive inference developed or not, Trabasso's model did not really explain how we perform transitive inferences. He proposed an alternative mechanism that he claimed did not require transitive inferences, which he named the **Sequential-Contiguity Model.**

Breslow proposed that subjects can understand the relational terms only in a categorical fashion. The midterms will all be ascribed contradictory attributes so the end-points are learned first. When a child learns B>C and B<A categorical labelling will fail but the child will at least come to retain the fact that B goes with A and C. This learning of contiguity allows the subject in some manner to form a linear order.

Whereas the Trabasso model stipulates that the child can only build the linear order once they have understood the premises as ordered pairs (A, B), the sequential-contiguity model specifies on the contrary, that children form the order on the basis of unordered pairs, e.g. (A, B) or (B, A), on the basis of contiguity relations.

Thus the double comparatives, used in Trabasso's version of the task, facilitate premise training by inhibiting categorical size labelling and thus increasing the young subjects reliance on contiguity relations. However, Breslow does not state how contiguity relations can give rise to the linear order. If a child knows that

B somehow goes with A and C, that C somehow goes with B and D, and that D somehow goes with C and E, but knows nothing more about B, C or D there is no way that the child can perform correctly on the BD comparison. Breslow talks about the series being built ends inward but is vague about exactly how this is done. It seems as if some form of directionality is required. If, however, we incorporate some form of directionality into Breslow's model, we end up with a model that is not noticeably different from Trabasso's.

At least in the sequential-contiguity model the linear order is postulated as the product of learning - not a pre-existent representational device into which everything must be squeezed whether it fits or not. Given that transforming the information in the premises into a linear order is a data reducing exercise, it may well be the case that contiguity relations play some part. This view is going beyond Breslow's, however.

Whatever, the status of the linear order, there is no way that it can account for the drop in performance to triads. However, it is worth noticing that in the dense triadic phase, the contiguity relations do change, so there is a possibility that this change might be the cause of the self-reparation.

2.4 The Stack Model

Harris and McGonigle, (Harris 1988, Harris and McGonigle 1993), developed the **Stack Model** to account for the monkey data. The **Stack Model** involves a simple production system. The strategy for performing the task consists of having a small set of rules, each one of which is an instruction to either avoid or select a particular colour. Along with these rules is a simple control strategy.

For example, for the series $A > B > C > D > E$, a possible stack might be:

1. if A is present then select A.
2. if E is present then avoid E.
3. if B is present then select B.
4. if C is present then select C.

The interpreter is as follows:

1. Remove a rule from the top of the stack.
2. If the rule is not applicable to the trial then go back to step 1.
3. Carry out the action of the rule and stop.

Given five items and two sorts of rule, the number of different stacks that contain four rules is 10 x 9 x 8 x 7 = 5040. Of these there are 16 (2^4) variants that perform correctly on the adjacent pairs. The most important fact however, is that it is a property of this mathematical space that any stack which performs correctly on the adjacent pairs also performs correctly on the remote ones. It is this property of the mathematical space that is the reason the model can successfully perform transitive inferences.

This model not only captured the transitive bias in a completely different way to other models but also by relating reaction times to the depth of search of the stack, the model also captured the distance effect. Harris and McGonigle further managed to assign individual subjects to specific possible stacks on the basis of reaction time and choice profiles. This was the first model to account for individual differences as well as the grouped data. Unfortunately, they were unable to test any predictions about the specific individuals based on the assignments made.

The **Stack Model** does account for the loss of performance on the initial triadic transfer tests, (the first model to do so), but it cannot account for how the subjects manage to repair their performance. The main reason for this is that although Harris did develop an inductive mechanism for learning the stacks, it was the last addition to the model and does not really contribute anything other than an existence proof that it is possible to learn stacks of rules. Nor could the inductive mechanism account for the differences from different training schedules.

The **Stack Model** is unable to cope with a circular relation. This does not mean that the Stack Model is no good – firstly, no monkeys have been tested on a circular relation yet, and secondly, there is always the possibility that the monkeys use a stack of rules for transitive relations and something else for intransitive ones. Of course, this makes the model less general and leaves unanswered the question of how the monkeys know which relations are transitive and which are not.

The **Stack Model** relates reaction times to the depth of search through the stack of rules. As this search takes place in a serial fashion it is very easy to make predictions about reaction times. However, Harris and McGonigle(1993), conclude at the end of the paper that the model should be changed such that the rules are activated in parallel, thus moving more towards a connectionist model and away from a production rule model. Obviously one effect from this would be to lose the connection between depth of search and reaction times.

2.5 The Back-Propagation Model

St Johnston(1989) modelled the transitivity task used by Trabasso and Riley(1975) in a back-propagation connectionist network. There were seven input units corresponding to the five different coloured rods and a unit for each relational term, *bigger-than* and *smaller-than*. There were five hidden units and five output units corresponding to choosing a rod of a particular colour. Figure 2–2 shows how these units are connected. This was the first model to treat the learning of the premises as an essential component of the skill.

By modelling the transitive inference task in a connectionist framework St Johnston took as strong an interactional approach as it is possible to take. By "interactional", I mean that the representation that emerges over the hidden units is the result of the network architecture, the way the network impinges on the world (i.e. the choices it actually makes), and how the world impinges on the network (i.e. the feedback from the task). His model consisted of a network of layers of units with no connections between units in a layer but with all units fully connected to the units in adjacent layers. The learning algorithm used was the back-propagation learning algorithm (Rumelhart, Hinton and Williams; 1986). With this algorithm the net was exposed to an input pattern corresponding to the encoding of the stimulus on a specific trial.

The network learnt the premises in a similar number of trials to the subjects. During learning there was a clear serial position effect. It generalised correctly to

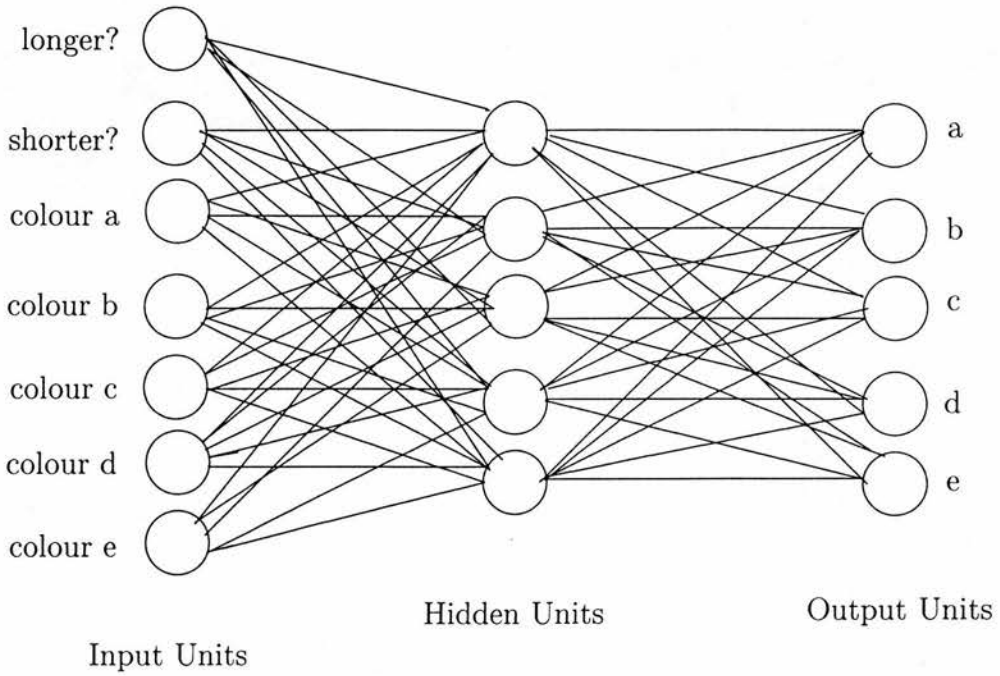


Figure 2–2: The Network for the Back-Propagation Model

the remote pairs, and interpreting the level of unambiguity as inversely proportional to reaction time there was a clear distance effect. When given a circular set of premises, the model not only had no problem in learning them, but also generalised according to the principle deemed most sensible: according to the shortest distance around the circle.

So as a direct result of treating learning as a central component of the task, we end up with a model in which a transitive bias is a default but not an absolute constraint, (shown by the ability to cope with a circular relation). The serial position effect and the distance effect were present despite the fact that there was no serial representational device. In fact, unlike the **Stack Model** where the rules are strictly ordered, there is not a hint of seriality anywhere to be seen - everything is parallel.

Despite these promising features, there are severe drawbacks to the model. Firstly, the model was completely impervious to different training schedules. It learns equally well whether the order is serial or random. Secondly, the model generalised perfectly to the triads.

Thirdly and most damningly, the model only achieved what it did because with the bidirectional version of the task and the representation used the mapping from inputs to outputs becomes linearly inseparable. That is to say that only with the bidirectional version of the task, is a hidden layer of units and a non-linear activation rule required. Thus, the network is forced to recode the representation over the hidden units. It is this forced recoding that produces all the effects. It is unreasonable, however, to assume that this is what is happening in subjects. Although using both comparatives seems to facilitate the task, it is by no means necessary to produce transitive inferences, and if only one comparative is used then there is no need for the hidden units. Furthermore, the number of hidden units greatly affects the sort of recoding that occurs - too many and there is no useful generalisation and too few and the network cannot learn the premises. This makes the model very task specific and ungeneralisable.

Fourthly, back-propagation networks are not self-supervised – they require target patterns. The transitive inference task is a forced choice task; i.e. subjects

have to choose one of the colours presented. If a subject is presented with a green stimulus and a red stimulus on a particular trial, then the subject cannot pick a yellow stimulus. Thus, the rewards given are not the only information available to the subject – there is a lot of implicit information as well. In order to get a transitive bias St Johnston had to experiment as to how the model should use this implicit information. If the model ignores it completely, there is no transitive bias. If the model uses treats the implicit information identically to the explicit information then there is no transitive bias. In order to get a transitive bias St Johnston used a criterion for using the implicit information in learning which he dubbed *pertinance*: if the activation for a non-presented colour is greater than the activation for the correct colour, then the fact that the preferred colour is not present becomes pertinant and is used in learning, otherwise activations for non-presented colours are not used in learning. This hypothesis on how subjects use implicit information could be tested empirically by changing the task such that the colours of the presented stimuli are separated from the available colours to choose from; e.g. present colours a and b but allow the subject to choose from a, b, c, d or e. The data from the triads suggests that this would have an effect, but the experiment has not been done. The main problem with the pertinance criterion is that it greatly increases the cognitive complexity of the model, but was largely forced on the modeller by the type of connectionist network used. Back-propagation networks are not self-supervised and need a target pattern for learning, and the pertinance criterion introduces some self-supervision but does not specify why it should be necessary or how it might be implemented within the framework.

2.6 Value Models

2.6.1 Value Transfer Theory

Ferson, Wynne, Delius and Staddon(1991) proposed a modified reinforcement theory which they dubbed the **Value Transfer Theory**.

According to Value Transfer Theory the subject assigns each stimulus a value. The value is made up of two components: the direct component which is the value that is conferred on a stimulus by the reward received for choosing that stimulus, and the indirect or transferred component which arises through partial generalisation from the values of the stimuli presented with a particular stimulus during learning. The subject chooses a stimulus according to the relative values of the two stimuli presented.

This idea is very similar to Breslow's Sequential Contiguity Model, in that transitive inference is the result of partial generalisation arising from the fact that different stimuli are presented together. Whereas, Breslow was vague about exactly what transfer takes place, Fersen, Wynne, Delius and Staddon provided some weird and wonderful value transfer functions. This is a prime example of modelling data rather than trying to model the subject – they give no account of why any of these functions should exist, except, of course, that they make the model fit the data!

2.6.2 The Honeybee Conditioning Model

Couvillon and Bitterman(1992), attacked **Value Transfer Theory** on the basis that there are much simpler ways of producing transitive inference. They provided a simple conditioning model of the pigeon data, based on a model of choice in honeybees. Similarly to the **Value Transfer Theory**, their models learned values for the different colours that were then used to determine choices. In order for this model to produce a strong transitive bias two *ad hoc* requirements were needed. Firstly, the learning rate was about ten times greater for incorrect responses than for correct ones. Given the fact that all of these types of model are bound to produce a serial position effect, this difference causes a weak transitive bias. Secondly, they chose an exponential choice function such that any small bias would be magnified. They give no explanation for why an agent should learn more from negative reinforcement than positive reinforcement, nor why the choice function should be as it is. There may well be good reasons for both of these but

until we know what they are, we must treat this conditioning model as a hack that provides little insight.

In the pigeon experiment the subjects were given correction trials when they made a mistake. Couvillon and Bitterman thought this was a necessary to produce a transitive bias, (even though no monkeys ever got the correction procedure!). Wynne, Ferson and Staddon(1992), however, showed that the model produces the transitive bias whether the correction trials are included or not. They then tried to reduce the free parameters, such as removing the exponential choice function and making the learning rates the same. They found that as you remove these the fit with the experimental data gets progressively worse. I do not think they surprised anyone there!

2.6.3 Problems with Value Models

Both the above two models assigned values to the different stimuli. Both sets of modellers also tried to distance themselves from 'mentalistic' linear representation devices. However, I cannot see any important difference between them. When you assign different objects different values, you are effectively ordering the objects along a one dimensional axis. In other words you are placing them into a linear array. Apart from pointing out that the modellers do not really know what they are trying to do, this also reveals that the models face exactly the same problems as the linear representational device model: namely, that there is no way they can cope either with the imperfect transfer to triads, nor the ability of pigeons (shown by the first set of modellers) to learn a circular set of premises having already learned a series. It is not fair to criticise too much for not being able to cope with the triads because the pigeons have never been tested on them, and we do not know how they would perform; (Ferson et al were not shy about pointing out the similarities between the two species, however, such as the serial position effect, and the symbolic distance effect – although they had not realised that the symbolic distance effect is actually based on reaction times not choice data!).

Neither the **Value Transfer Theory** nor the **Honeybee Conditioning Model** have been able to account for the difference in training schedules either. In fact the only thing recommending these two models is that they treated learning as an integral part of the acquisition of the transitive bias; the rest is completely uninteresting. It should be pretty obvious by now that not everything can be reduced to conditioning, yet that is what seems to motivate these researchers. Of course parsimony should be a goal for any modeller, but parsimony at the sake of everything else is pointless. Afterall, the debate as to whether conditioning theory is sufficient to account for all behaviour, was settled long ago with the Behaviourists losing.

Chapter 3

Self-Supervised Learning Models of Transitive Inference

3.1 Modelling Strategy

The aim of this modelling is to capture performance on a transitive inference task in a 'Piagetian' way. That is to say, that the models should self-regulate themselves through interaction with the task such that they learn the premises and generalise automatically to unlearnt stimuli combinations. When the relation being learnt is a transitive one then we should expect some implicit ordering over the items, and when the relation is intransitive (i.e. circular) the model should still be able to cope.

3.1.1 Choosing a Modelling Framework: Temporal Difference Learning Models

Sutton and Barto(1987,1989), have developed a framework bringing together ideas from animal learning (e.g. conditioning), and synthetic learning systems (e.g. stochastic dynamic programming), to account for learning and sequential decision-making in the animal psychology literature. Although the transitive inference task is not, *logically speaking*, a sequencial decision-making task, (i.e. a decision made on one trial does not affect what decision should be made on the next), the fact that Kallio(1982), found that different training schedules can have an effect suggests that maybe the subjects do not see the task from the logicians point of view.

Stimulus Pairs	a	b	b	c	c	d	d	e
Rewards	+	-	+	-	+	-	+	-

\+ peanut

\- no peanut

Table 3–1: Stimulus Pairs and Rewards.

More importantly, this sort of model is self-supervised. The model is not given a target to try and reproduce, as with, say, back-propagation learning networks, but must learn to evaluate its own performance and then change its policy for behaving accordingly. Thus, this type of model can capture learning and generalisation effects.

3.1.2 The 5-Term Series Task

The version of the task modelled here is essentially that used by McGonigle and Chalmers(1984).

There are five tins each of a different colour. The actual colours used are not important to the modelling; here they are simply labelled a, b, c, d and e. On each trial during the training phase, two of the tins are presented. One tin has a weight in it making it heavier. Both tins are placed on a tray in positions underneath which are small cavities. For half the subjects there is a peanut always under the heavy tin, and for the other half there is a peanut always under the light tin. The subjects only get to displace one tin per trial. Obviously, the subjects cannot *see* which tin is heavier, so they have to learn which of the pair contains the weight and hides the peanut, whilst interacting with the task. The subjects are trained on the four premises ab, bc, cd and de, and then tested on the remote pairs ac, bd, ce, ad, be and ae. The training pairs and rewards are shown in Table 3–1.

The premises are "taught" to the subjects in several progressive training schedules. First, the subjects are trained on each of the training pairs until they reach a criterion of performance. For this first stage, first the pair ab is taught, then bc, then cd and finally de. The next stage of training involves giving each premise for

four trials in succession before the next premise in the series. This is reduced to two in succession and then to one, which forms what has been dubbed the serial schedule. In the serial schedule the premise *ab* is presented on the first trial, the premise *bc* on the second, *cd* on the third and *de* on the fourth. The fifth is *ab* again, and so it carries on.

Once the subjects have reached the criterion of performance on the serial schedule they are moved on to the final stage of training, the random schedule. In the random schedule the four premises are each presented to the subject before another premise is shown twice, but the order that the four are presented is random. One of the experimental effects we are trying to capture is the difference in the difficulty learning the training pairs when the schedule is random and when the schedule is serial. For the purposes of modelling then, the models presented here were given the two pure schedules: serial and random.

The test phase for the binary pairs consists of presenting all ten possible binary pairs in random order and giving a reward for whichever tin is chosen; that is, there is no discriminative feedback. With the model it is possible to see what the choice would be likely to be on a test pair without actually presenting it. This way it is possible to get a much richer view of how the transitive bias arises, (if it does), without interfering with the training procedure.

Additionally to the binary test pairs the subjects were also given all six combinations of three colours of tins, (triads). No discriminative feedback was given for the triadic trials. At first the triadic transfer tests were "shallow" as with the binary test pairs. That is the trials were embedded with the binary training pairs and given only enough to test the competence of the subjects without giving them enough time to significantly alter their performance to the new trials. Later the triads were given in a "dense" phase, where the subjects still received no discriminative feedback but did get significant experience with the new type of trial.

Although, McGonigle and Chalmers never gave a circular set of premises to the monkeys, it is important to show that a transitive bias is not somehow built into the models, which would beg the main question. If a transitive bias was built

in, then it should be impossible for the models to perform correctly on a circular set of premises.

3.1.3 The Behaviour the Models should Capture

The monkeys can learn the premises of the five-term serial task given that they are presented them often enough. During learning they show a serial position effect, such that the premises involving end-points are learned fastest, and the premise in the middle but nearer the top end of the scale is learned faster than the one nearer the lower end of the scale. The monkeys were never given pure training schedules from scratch, but as these models are an attempt to understand performance in terms of learning and self-regulating, I am interested in capturing Kallio's result that the premises are much easier to learn when presented in a serial schedule then when presented in a random one.

The monkeys show a strong transitive bias on all remote pairs. Their reaction times show a strong distance effect such that the further apart the two items to be chosen from, the faster they respond. The models here are instantiated as connectionist networks. These give clear choice profiles but do not generate reaction times without further assumptions which are difficult to test or justify. Therefore, although it would be nice to capture the distance effect as it has played such a central part in the transitive inference literature, these models must remain silent on this experimental effect.

On the triadic transfer tests, the monkeys initially show a drop in the transitive bias, but when the triads are presented in the dense transfer phase, some subjects repair their performance even without discriminatory feedback. The subjects that do not repair spontaneously only require a minimal number of trials with feedback to do so.

Ferson, Wynne, Delius and Staddon(1991), showed that pigeons could learn a circular relation, having learnt a serial one made from the same items, even though they generalised with a strong transitive bias. I make the assumption that if pigeons can do it, then so can squirrel monkeys.

This catalogue of behaviour that I want the models to capture can be summarised as a set of questions.

1. Can the model learn the premises?

2. Does the model show a serial position effect?

3. Is the model sensitive to training schedules?

4. Does the model show a transitive bias?

5. How does the model respond to triads?

   (a) Does it perform perfectly?

   (b) Does it perform imperfectly but recover spontaneously?

   (c) Does it perform imperfectly but cannot recover whatever?

   (d) Does it perform imperfectly but can recover only with differential feedback?

6. Can the model cope with a circular relation?

   (a) When taught from scratch?

   (b) Having already learnt a linear one?

7. Does the model generalise appropriately on a circular relation?

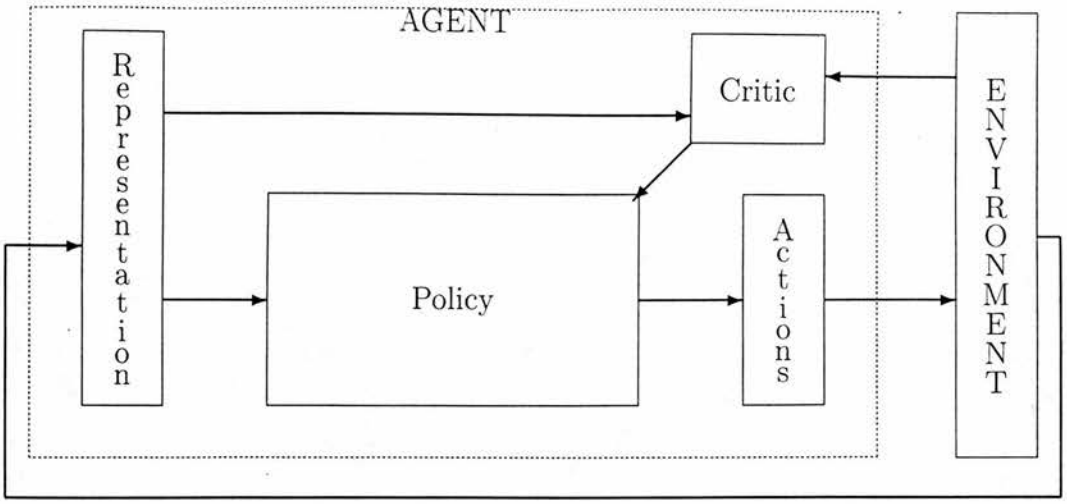


Figure 3–1: Components of the Basic Model and How They Fit Together.

3.2 Basic Components of the Models

There are five basic components: an environment, a representation of the environment, a set of actions to perform on the environment, a policy for choosing actions, and a critic to evaluate the policy. Figure 3–1 shows the basic components of the model and how they fit together.

The Environment The environment provides information for the agent and delivers rewards for actions that the agent performs. The environment need not be external to the organism; it could, for instance, consist at least partly of internal memory. From this it should be noted that 'agent' does not necessarily refer to a whole organism, just to a specific decision making module within an organism. In modelling this task, however, the environment does consist of all those things which are external to the agent. Thus the environment includes the coloured tins, the tray, the peanuts and everything else that might impinge on the monkeys sensors. In essence, the environment embodies the task.

A Representation The agent is not assumed to be omniscient. In fact, a large part of this modelling methodology involves experimenting with just how much information the agent can pick up from the environment. This is done through the representation. The representation encodes salient features from the environment that can be used to inform the decision making process. 'Representation' is used in a technical sense here; each feature is best considered as an experiment done on the environment which determines whether the specific feature should be on or off.

A Set of Actions The agent impinges on the environment by performing an action. In this task the only actions considered are displacing a tin in order to retrieve a peanut that may or may not be hidden underneath it. There is a specific action associated with displacing a tin of a specific colour. Only a subset of the tins are present for any one trial, so the set of actions is reduced to a set of possible actions from which one must be chosen in each trial.

The Policy The agent has a representation of the state of the environment and a set of possible actions from which to choose; the policy provides the mapping from the represented state of the environment to the set of possible actions. This mapping need not be deterministic. In the models presented here, there is a deterministic component and a stochastic one. The policy determines a level of support for each of the possible actions. A random number is then drawn from an exponential distribution for each of the possible actions which is added to the level of support for that action. The possible action with the highest value is then chosen as the action to perform at that moment. The purpose of drawing the random numbers from an exponential distribution is to make small numbers more likely to come up than high numbers. The reason for having a stochastic element at all is because it is necessary to have some exploration embodied in the policy in order to allow it improve itself effectively.

The Critic The agent needs to evaluate how good it's policy is, and in most cases the feedback it receives from the environment is extremely poor. The

critic instantiates an evaluation function for the policy, which for each state predicts the expected return from the environment which it can then compare with what is actually received and thus produce a measure of how well the policy is performing. The critic does this without knowing anything about what the policy is doing, how it does it, or what particular action was taken.

At each time step the representation picks up information about the current state of the environment. This information is fed into the policy which decides which action to perform, and into the critic which predicts the expected reward that the agent will receive for that action. The action performed changes the state of the environment and the critic determines how the expected reward compares with the actual reward. The critic uses this comparison to change the policy and to change its future expectations. The clock then ticks on one time step, the representation updates it's information on the state of the environment and another action is chosen and evaluated.

3.3 Applying the framework to the 5-term series transitive inference problem

Representation

For each colour there is an input node which is on if that colour of tin is present and off if it is not. Thus the four premises are encoded in the following input vectors:

ab 11000

bc 01100

cd 00110

de 00011

The first dimension of the input vector is assumed to correspond to the tin of colour a, the colour at the top of the series, the second to b, and so on down to e, the colour at the bottom of the series.

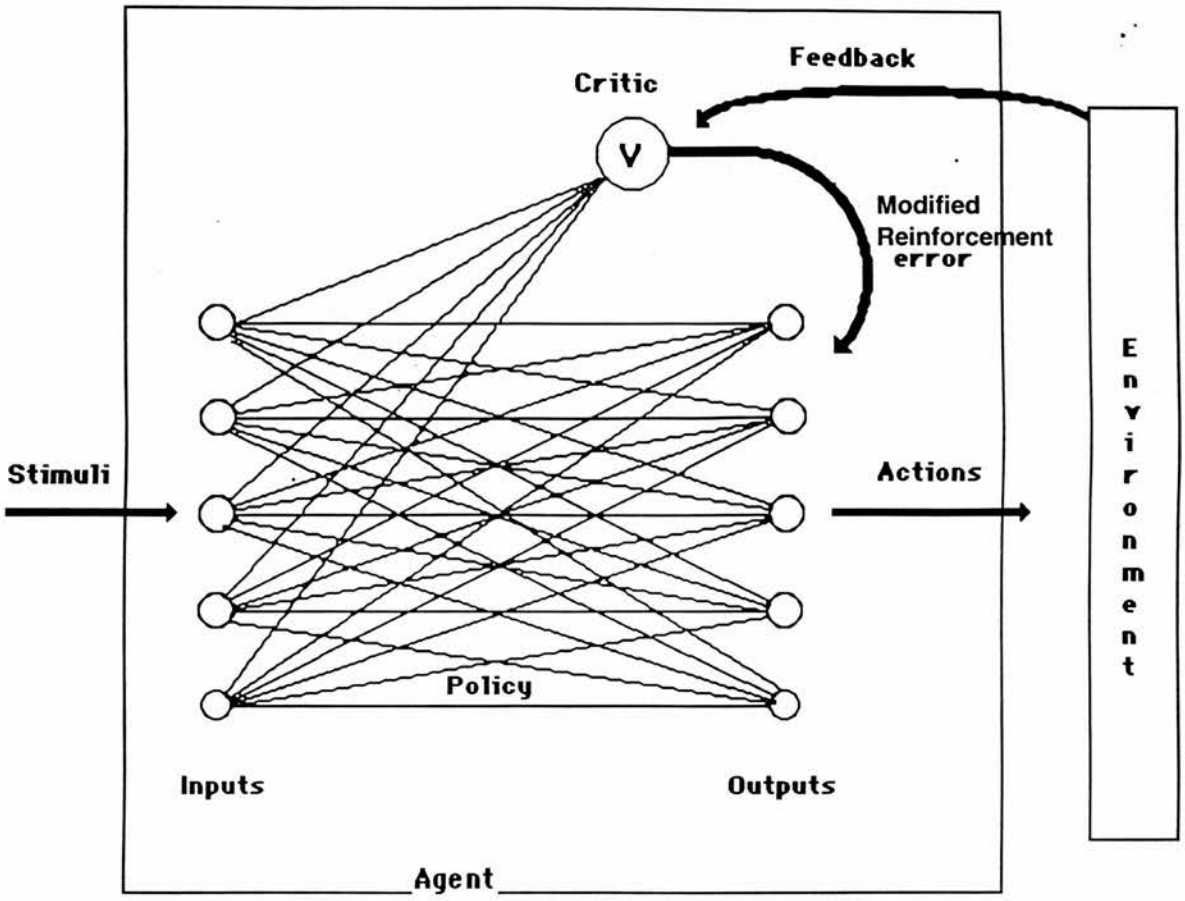


Figure 3–2: Connectionist Instantiation of the Basic Model.

For each colour of tin there is an action which corresponds to choosing that tin. The action of picking tin a is denoted as A, picking tin b is denoted as B and so on. The correct action to choose is the one nearest the beginning of the alphabet of those presented.

3.3.1 Instantiating the agent as a connectionist network

Figure 3–2 shows how the components that comprise the agent are naturally implemented as a connectionist network. The representation forms a set of input nodes whose activations at each time step t form a vector, $\Phi(x_t)$ representing the environment at that time step. Each action a in the set of actions is a node in

the output layer. The policy, π, is simply the weight matrix produced by fully connecting the input layer to the output layer. The activation for each possible action is produced by multiplying the input vector $\Phi(x_t)$ by the weight vector $\pi(j_t)$ for the corresponding action a_j, (a straightforward linear activation rule). The activation for each action a_j is then increased according to the stochastic part of the policy explained earlier.

Expected Discounted Return

In some tasks the feedback is at an absolute minimum. Most of the responses get exactly the same feedback as each other, and occasionally a response will finish the trial and a peanut will be dispensed. The agent needs to be able to evaluate responses in a much more sophisticated way than simply as to whether a particular response received a peanut. A measure that might be used to evaluate responses, and that would be more useful, would be a measure of how far (in time) the agent is from positive reward, (even if it were only a rough estimate). One way of doing this is by using *future discounted rewards.* The total discounted reward from time 0 is defined to be

$$\sum_{t=0}^{\infty} \gamma^n r_{n+1} = r_1 + \gamma r_2 + \gamma^2 r_3 + ... + \gamma^n r_{n+1} + ...$$

where r_t is the reward the agent receives at time t and γ is the discount factor. γ is between 0 and 1, so discounted reward gives greater weight to rewards that are received sooner than those received later. Of course, the agent does not know what r_t is for t into the future, but it is relatively easy to learn to compute the expected value of the discounted reward.

The Transitive Inference task is not a sequential task, i.e. the decision on one trial is not affected by the decision on the previous one, but by assuming that the agent is geared up to solve serial tasks we may find the sequential characteristics of the subjects in this non-sequential task.

Learning an evaluation function

The critic learns to predict expected discounted reward and is instantiated as a unit which receives inputs from the input layer and outputs a number which is the current estimate for the expected discounted return for the current policy. There is a vector v such that $v^T\Phi(x) = V^\pi(x)$; i.e. the expected discounted return is computed as the weighted sum of each dimension of the input vector.

The evaluation is learned by comparing predictions on two adjacent time steps. Hence, the name Temporal Difference Learning. The error, ϵ, is computed on the following time step:

$$\epsilon_{t+1} = r_{t+1} + \gamma V_t(x_{t+1}) - V_t(x_t)$$

where r_{t+1} is the reward received for the action taken at time t, $V_t(x_{t+1})$ is the expected future payoff, and $V_t(x_t)$ is the future payoff that we had expected before taking action, a_t. It is easy to see that $r_{t+1} + \gamma V_t(x_{t+1})$ and $V_t(x_t)$ are both approximations of the same quantity, the expected discounted return from time step t, but the former is based on more upto date information. The error is just the difference between the new estimate and the old estimate.

The vector, v, that determines V, is updated using one of the variants of the delta rule:

$$v_{t+1} = v_t + \beta\epsilon_{t+1}\Phi(x_t)$$

thus:

$$v_{t+1} = v_t + \beta[r_{t+1} + \gamma V_t(x_{t+1}) - V_t(x_t)]\Phi(x_t)$$

where β is the evaluation function learning rate. So the evaluation function is learned using a standard gradient-decent mechanism. Notice, that it learns the evaluation function for the current policy. If the policy changes then the evaluation function should also change. Within this framework it is assumed that the policy is being learned at the same time as the evaluation function, thus the adjustments to the policy may not improve it at each time step because the adjustments depend on the current evaluation function which is not necessarily the optimal evaluation function for the current policy. However, the evaluation function will tend to improve over time and hopefully the policy will also tend to improve.

Improving the policy

The policy is improved using an operant conditioning mechanism. There is a separate parametized function, π_j, for each action, a_j. Each of these functions assigns a number to a state. An action is then selected by comparing the numbers produced by these functions for the current state. The action chosen is the action that wins a stochastic competition. Learning only occurs for the function of the action that is chosen but because of the competitive mechanism this can effect the liklihood of other actions being taken.

The basic idea of operant conditioning is that actions that lead to greater rewards should be made more likely to occur in the future and actions that lead to lesser rewards made less likely. The problem that needs to be solved is to find a way of telling whether a specific reward is greater than average or not. The specific reward an action gets is $r_{t+1} + \gamma V_t(x_{t+1})$. It so happens in this formulation we already have a measure of average reward: $V_t(x_t)$. To see why this works consider the error function for the evaluation policy. As the error tends to zero $V_t(x_t)$ should approach the average of $r_{t+1} + \gamma V_t(x_{t+1})$. Thus, we can use the same error to update the policy as we do to update the evaluation function.

3.3.2 Parameter Settings

There are a number of parameters that must be set for each run of the model. All of the models presented in this chapter need the following five parameters: value of a correct answer ($R+$), value of an incorrect answer ($R-$), policy learning rate (ρ), evaluation function learning rate (β), and the discount factor (γ). For all the runs of all the models presented here the values of $R+$ and $R-$ are set at 1 and 0 respectively. These values are intuitive but the only constraint on these values is that $R+$ is greater than $R-$. The initial values of all the weights in the policy and evaluation function start at zero. Therefore, the initial conditions taken with the value for $R-$, assumes that the model has no initial colour preferences and expects no reward for any choice. This seems as reasonable a characterisation of

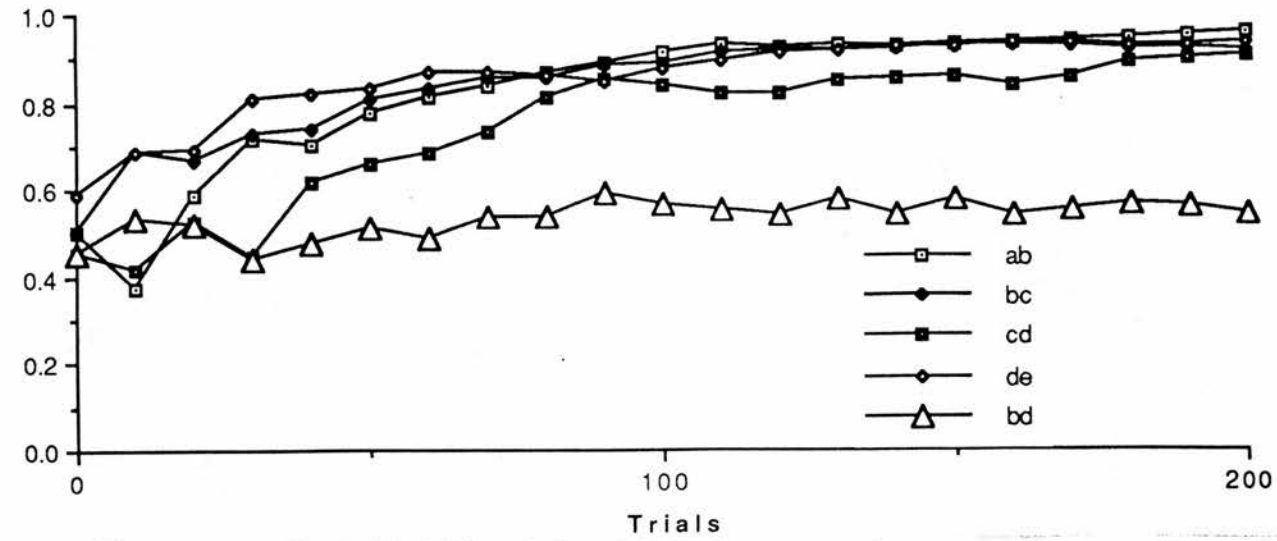


Figure 3–3: Basic Model:Serial Schedule with $\gamma = 0.9$, $\beta = 0.3$ and $\rho = 0.1$.

the initial state of the actual subjects as any, but any of these assumptions can be relaxed to a moderate extent without significant effect.

The purpose of this modelling was not to find some optimal set of parameter values such that a relation can be learned as fast as possible. In fact, as stated earlier in this chapter, the questions I want to answer are exclusively to do with the qualitative performance not the quantitative performance of the model. This has led to a different type of exploration of the parameter space. The values of the parameters can affect the qualitative performance of the models and where they do so, I state it explicitly. On the whole, however, the models performance is remarkably robust with respect to the various parameter values. Therefore, the values chosen for the particular runs of the models are generally unimportant. If either of the learning rates is reduced the model learns slower but it still learns the same thing - it just takes longer. Where I have not explicitly stated otherwise the values used were chosen to produce the clearest results, given the particular question being asked.

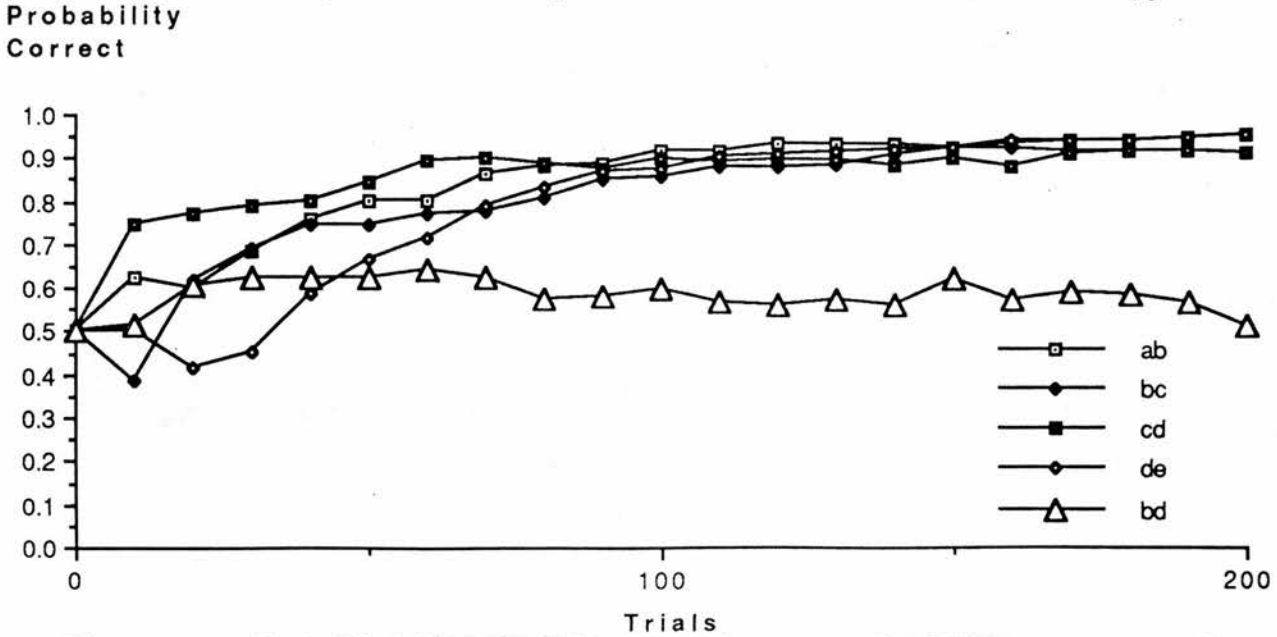


Figure 3–4: Basic Model:Random Schedule with $\gamma = 0.9$, $\beta = 0.3$ and $\rho = 0.1$.

3.4 The Basic Model

Figures 3–3 and 3–4 show typical results from the model for each of the two training schedules. The model never actually receives the test pairs; all the points on the graph show the probabilities of choosing the correct colour of tin *if* that particular pair of stimuli had been presented at that particular moment during training.

All the training pairs are learnt to a criterion of above a probability of 0.9 for choosing the correct colour, by 200 trials. However, the different training pairs are not learned equally fast. *ab* is learned the fastest with *de* not far behind. The training pairs not including end-points require considerably more training to reach criterion with *cd* tending to take the longest. Thus, there is clear evidence of an end-point and serial position effect despite the fact that the agent has no serial representational ability; it merely has a number of associations between colours being present and support for choosing colours.

The graphs show no evidence of a transitive bias, with the *bd* pair hovering around the 0.5 level. This is despite the fact that the models were run for considerably longer than was necessary for the premises to be learned.

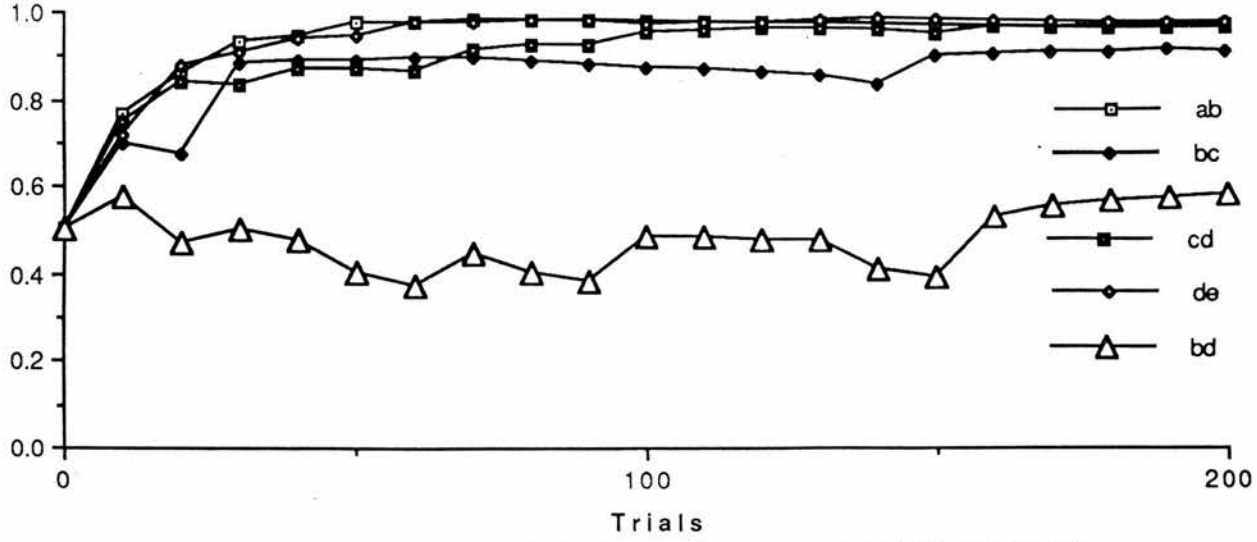


Figure 3–5: Basic Model:Serial Schedule with $\gamma = 0.0$, $\beta = 0.3$ and $\rho = 0.1$.

	Training and Test Pairs									
	ab	bc	cd	de	ac	bd	ce	ad	be	ae
serial $\gamma = 0.9$	0.97	0.93	0.92	0.95	0.82	0.55	0.77	0.80	0.80	0.91
random $\gamma = 0.9$	0.96	0.92	0.92	0.97	0.78	0.52	0.80	0.76	0.83	0.92
serial $\gamma = 0.0$	0.98	0.93	0.99	0.99	0.75	0.59	0.92	0.84	0.90	0.96

Table 3–2: Final Probabilities from the Basic Model.

Parameter Values

There are no important differences between the graphs for the two different training schedules despite the high value of γ. Figure 3–5 shows the learning curve for $\gamma = 0.0$. As can be seen there does not seem to be any effect of γ at all.

The model is extremely robust with respect to all the different parameters with one important exception: if γ is high and the policy learning parameter, ρ, is significantly high with respect to the evaluation function learning parameter, β, then the model can get into a 'depressed' state where it fails to learn at least one of the training pairs but just accepts the fact that it receives no peanut with this particular pair. Figure 3–6 shows an example of this happening.

Table 3–2 shows the final probabilities for all the binary pairs for the three runs shown in figures 3–3, 3–4 and 3–5.

Figure 3–6: Basic Model:Serial Schedule with $\gamma = 0.9$, $\beta = 0.1$ and $\rho = 0.3$.

Inputs					Actions
a	b	c	d	e	
+	+				A
−	+−	+			B
	−	+−	+		C
		−	+−	+	D
			−	−	E

Table 3–3: Grid Showing the Direction of Reinforcement to Particular Weights in the Policy.

Although the bias for the *bd* pair does not seem significant from the graphs, there is a very weak stochastic transitive bias. From twenty runs of the model sixteen had a weak transitive bias. The average bias from all twenty runs was 55%. The bias from the other test pairs was considerably stronger. This is the result of the strong end-point effect.

3.4.1 Examining the policy

Table 3–3 shows the types of reinforcement to each weight in the policy.

There is a column for each colour of tin presented, here labelled a, b, c, d, and e. There is a row for each action i.e. choosing a particular colour, here labelled A, B, C, D, and E. The symbols inside the grid represent the directions of reinforcement each weight in the policy receives. So, during training whenever tin a is present either A or B is chosen. When A is chosen the reinforcement will be positive and when B is chosen the reinforcement will be negative; hence the symbols $+$ and $-$ in the boxes corresponding to aA and aB respectively. The box bB has both a $+$ and a $-$ in it because when ab is presented choosing B gets negative reinforcement and when bc is presented choosing B gets positive reinforcement. Remember that only the action that is chosen gets any reinforcement but, because the choice is a result of a stochastic competition, changing the support for one action effects the probability of all the other actions being chosen.

End Point Effect

The grid shows clearly why an end point effect should arise. A and E are the only actions for which there is no ambiguity either within the boxes or between the boxes.

Weak Transitive Bias

For a transitive bias to arise the weight bB must be greater than the weight dD as the weights bD and dB will have remained at zero. If the weight bB increases because of a correct response to a bc pair then there will be pressure for aA and bA to become more positive and aB and bB to become more negative, but there will not be any pressure on the weights for action C. If on the other hand the weight dD increases then the weights cC and dC may have to increase and if cC does increase then there will be a pressure for bB and cB to increase. In other words increasing the weight dD gives rise to a pressure for the weight bB to increase but the reverse is not true. Conversely, if the weight dD drops it will only effect the de premise, whereas if the weight bB drops, it will create a small pressure for the weight cC to drop which in turn will affect dD.

This is the cause of the weak transitive bias. It is a much weaker bias than that found experimentally from the monkeys and children. If the net is overtrained to the extent that the probabilities of getting the premises right is over 0.99 for all four premises then the transitive bias might reach upto 70%. The monkeys showed a transitive bias of around 90% even when the premises were only learned to a similar level. When the net has learnt the premises to the 90% level, all that can be said is that B will be chosen just more than 50% of the time, on over half the runs; i.e. the bias is as weak as they come.

The **Basic Model** has a simple representation of the task and a self-supervised learning mechanism which is perfectly capable of learning the premises. However, it does not generalize to the remote pairs to anywhere near the same extent as the actual subjects being modelled. Therefore, the actual subjects must be doing something else or something different. If the actual subjects are doing something completely different then we need a completely different type of model. The strength of Sutton and Barto's self-supervised learning model, however, is that it makes so few assumptions about the subjects' knowledge of the task and what is expected. This is congruous with the types of subjects we are modelling (monkeys and pigeons). Furthermore, the learning of the premises is similar enough between the model and the subjects to suggest that this type of model in along the right lines. Therefore, it is not time to change the modelling framework, but time to investigate how to extend the model to explore what else is required by the model and used by the subjects to produce a strong transitive bias.

The performance of the **Basic Model** on triads corroborates this view.

Triadic Choices

The particular triadic choice patterns shown by the model vary from one run to the next; it depends on the history of actions chosen. It also depends on the values of the various parameters γ, β and ρ. However, the reinforcement table 3–3 does hint to general aspects of triadic choice. Each colour of tin supports choosing the colour immediately above in the rank $a-b-c-d-e$, and undermines the support

Choice	B	C	D
serial schedule, $\gamma = 0.9$	0.73	0.21	0.06
random schedule, $\gamma = 0.9$	0.68	0.26	0.06
serial schedule, $\gamma = 0.0$	0.54	0.44	0.02

Table 3–4: Performance on Triadic Transfer Test of Successful Runs.

for the colour immediately below it in the rank. Apart from this, the only other effects are the support for themselves. We know from the weak transitive bias that $aA > bB > cC > dD > eE$. Thus, for the crucial triad *bcd* we know that B will get some support from b, but most support from c, whilst C will get a little support from c, most support from d, and negative support from b. From this alone it is easy to see that the probability of choosing B for *bcd* will be considerably less than choosing B for *bc*, because for the latter there is no support for C from d.

Table 3–4 shows the results on the triad BCD for the three runs shown in the graphs and tables above for which training was successful.

The transitive bias is very weak so choosing B from *bcd* is still more likely than choosing B from *bd*. The model does show a potential to capture the triadic data, which a static linear array could never do.

3.4.2 Conclusions from the Basic Model

If the critic leads the policy or if the discount factor, γ, is not too big the agent has no problems learning the premises. There is no perceivable difference between the serial and the random training schedules, and apart from its role in learnability, γ seems to have no effect. There is a clear end-point effect and serial position effect and a weak transitive bias. The model captures the empirical finding that performance on the triad *bcd* is worse than performance on the training pairs *bc* and *cd*, but due to the weakness of the transitive bias, performance on this triad is still considerably better than performance on the *bd* test pair.

The **Basic Model** can be used as a baseline against which to measure the performances of extended models, where new competences have been added to the agent. We know that there is no built-in strong transitive bias in the general framework, just the weakest of weak stochastic biases. The questions now are how to change the model such that it produces a strong transitive bias and is sensitive to training schedules.

3.5 Prospective and Memory Units

Sutton and Pinette (1985) developed a means of allowing the agent to construct temporal contiguity relationships. Different training schedules by their very nature have different temporal contiguity relationships, so by incorporating units that predict these relationships into the existing model, we might uncover why the different training schedules give rise to such differences in competence.

These units, called *mapping* units [1], receive inputs from the each of the input units and output to the policy and the critic. There is one mapping unit for each of the input units. They learn in exactly the same way as the critic learns the evaluation function except the feedback is not the reward but the activation on the next time step for the corresponding input unit.

Mapping Units learn to predict the expected discounted future activation of each of the input units based on the current activation of those input units. More or less the same outcome can be achieved by looking backwards rather than forwards in time. An alternative to mapping units are memory units whose activations decay over time. Thus there is a memory unit for each colour of tin. The activation of these units starts at zero, When a particular colour of tin is presented at time, t, then the activation of the corresponding memory unit is incremented by 1 on the following time step, $t + 1$. This activation decays at each following timestep by being multiplied by a decay parameter, δ, whose value is between 0 and 1.

---

[1] Sutton and Pinette used recurrent units, but it makes no difference here.

The Memory and Mapping units did not capture the differences between the training schedules: both learned the premises in similar numbers of trials under both schedules. Surprisingly though, both produce strong transitive biases. Examining the policy showed that the aggregate of the weights between the new units and the actions provided a ranking of preferences in the order $A > B > C > D > E$.

These new units were not designed to give a strong transitive bias but to enrich the representation in a temporal way such that ambiguous states might be differentiated. There are no ambiguous states in this task, given the basic representation, and yet we do now get a strong transitive bias. One plausible explanation for the subjects' performance is that they elaborate their representations of the current state because they do not know, initially at least, what the task is and therefore what representational requirements are needed. This would imply that a strong transitive bias is merely a by-product of an inquisitive subject trying to understand the task and thus enriching their representation more than is necessary. Of course over-enriching the representation can lead to interference problems. The severity of these problems depends on the task and the specific elaborations but assuming that these extra features are unnecessary the most important factor will be how many extra features there are. This suggests that it that subjects would only elaborate their representations in a limited way. This then raises the question of what sorts of elaboratory features lead to a strong transitive bias. If the category is small then the chances of the strong bias being merely a by-product, (useful or not), are reduced. If on the other hand most elaboratory features lead to a strong transitive bias, then different questions are raised. Depending on the answer we should concentrate on, either how to achieve a strong bias, or how to cope with it.

Strength of Bias	Training and Test pairs									
	ab	bc	cd	de	ac	bd	ce	ad	be	ae
0	0.98	0.93	0.99	0.99	0.75	0.59	0.92	0.84	0.90	0.96
1	0.97	0.96	0.90	0.97	0.92	0.71	0.86	0.91	0.92	0.98
2	0.98	0.94	0.92	0.97	0.95	0.80	0.92	0.97	0.95	0.99
3	0.92	0.96	0.95	0.98	0.93	0.93	0.97	0.97	0.99	1.00

Table 3–5: Final Probabilities from the Bias Model with Different Strengths of Bias.

3.6 The Bias Model: Constant Elaboration

Is it necessary to have specific elaboratory units to produce a strong transitive bias? The answer is no. A weak transitive bias is produced without any elaboratory units whatsoever, and a weak transitive bias can be turned into a strong transitive bias simply by amplifying it. The **Bias Model** is the same as the **Basic Model**, except that it has a bias unit in the input layer which always has an activation of one. The bias unit is always on so it cannot be used in a discriminatory way. Therefore all it can do is give a bias to each of the possible actions, (hence its name). In fact having one bias unit is exactly equivalent to doubling the learning rate on the weights between a stimulus being present and choosing that stimulus. Having two bias units, or having one with double the learning rate, is equivalent to tripling the learning rate of the xX weights.

3.6.1 Performance of the Bias Model

The parameter values were exactly the same as for the **Basic Model**. Table 3–5 shows the effects of the bias units. The strength of the bias in the table is best understood as the number of bias units added to the model, so the first row of table 3–5 is the same as the **Basic Model** and just included here for comparison.

Choice	B	C	D
Bias 1	0.80	0.17	0.03
Bias 2	0.90	0.08	0.02
Bias 3	0.90	0.10	0.00

Table 3–6: Performance on the *bcd* Triad for the Bias Model.

I explained in the section on the **Basic Model** how there is a pressure for a small transitive bias to emerge. With the Bias units this is amplified. The greater the number of bias units the greater the amplification. Another way of understanding this pressure is to realise that a transitive bias instantiated in this way actively helps the learning of the premises. Therefore, if the network can represent a transitive bias then there is always going to be a tendency to do so. The policy has a number of degrees of freedom in which to manoevre. Exploration of this multidimensional space progresses during training. The bias units accentuate one particular degree of freedom from the **Basic Model**. All the bias units can do is to add a preference ranking to picking a particular colour. If this preference ranking is not $A > B > C > D > E$ then the performance on at least one of the training pairs will be interfered with by the bias units. Any incorrect response will then lead to the preference rankings moving towards the ranking mentioned above.

3.6.2 Triadic Transfer

Table 3–6 shows how the **Bias Model** performed on the triads.

As the bias increases the transfer to the triads increases. So this model cannot explain why the monkeys' performance deteriorates. The **Bias Model** cannot produce a strong transitive bias and at the same time show significant deterioration in triadic transfer tests.

3.6.3 Conclusions from the Bias Model

It is worthwhile here comparing the **Bias Model** to Couvillon and Bitterman's **Honeybee Conditioning Model**. Both generate a small transitive bias and then amplify it. The ways that they do these things are different however and the conclusions drawn must be different accordingly. The **Honeybee Conditioning Model** produces a small transitive bias by making use of the experimental fact that some of the premises are learned slower than others, (the serial position effect). The serial position effect is converted into a weak transitive bias by having different learning rates for correct and incorrect responses. In the **Bias Model**, as in the **Basic Model**, the small transitive bias is totally inherent to the way the model represents the premises. The **Honeybee Conditioning Model** amplifies this bias by using an arbitrary choice function. With the **Bias Model** the amplification is not hidden in an *ad hoc* mechanism, but openly stated and instantiated in terms of the simplest possible elaboration to the model. Couvillon and Bitterman conclude that they have come up with the most parsimonious model of transitivity. Even ignoring the arbitrary elements in their model which they have tried to hide, their model may be parsimonious, but it is also particularly uninteresting when you consider the insights it gives and the possible ways to extend it, (there do not seem to be any).

With the **Bias Model**, on the other hand, we can conclude that a transitive bias is incredibly easy to achieve: all you have to do is elaborate the input layer with other units; even random noise would do! Bias units are the most simple and parsimonious but they are by no means the most interesting. The exciting thing now is to think of more useful units to elaborate the representation with, that not only produce a bias, but also allow the agent to solve different tasks – tasks where a transitive bias might be a hinderance not a help.

3.7 The Contiguity Model

Breslow suggested that when the subject learns to choose B over C that somehow the fact that the agent has already learned to choose A over B is encoded somehow. That is to say that learning $bc \rightarrow B\&\neg C$ produces something of $ab \rightarrow B\&\neg C$.

If the **Sequential-Contiguity Model** is independent of temporal sequence then the contiguity must be in terms of overlap between ab and bc. We know from the first premise that $a \rightarrow \neg B$ so the overlap must induce $a \rightarrow \neg C$.

To do this we need to get the agent to represent the fact that a is presented with b half the time. We can achieve the effects we want by elaborating the inputs with units the learn stimulus - stimulus relationships. One way of doing this is by having another set of units that try and reproduce the pattern presented based on how the components of the inputs support each other. Figure 3–7 shows the connectionist instantiation of this.

In this model there are five extra units. They receive activation from the input units and they output to the policy and the critic. The weights between the inputs and the contiguity units start at zero and as training progresses the units learn that a is only presented with b, that b is presented half the time with a, and half the time with c and so on. We cannot assume that the subjects know these values by magic, but we can get the agent to learn them. Each contiguity unit is connected to every input in the input layer.

The error function for these units is

$$\epsilon_i = I_i - C_i$$

where C_i is the activation of the Contiguity unit corresponding to the stimulus i, and I_i is the activation of the input unit. The update rule is given by the rule

$$\Delta w_{ij} = \epsilon_i . I_i . I_j . \sigma$$

where w_{ij} is the weight connecting the input j to the Contiguity unit i, and σ is the learning rate.

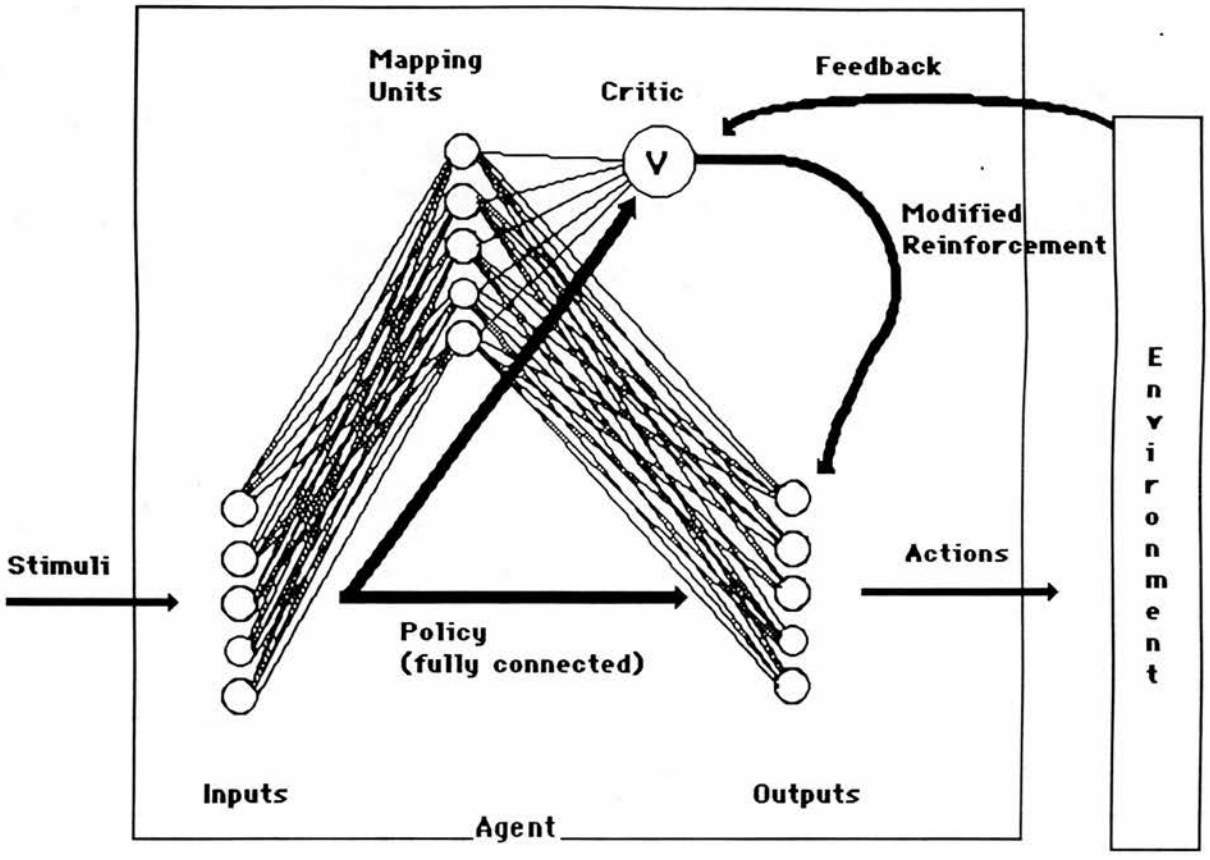


Figure 3–7: Connectionist Instantiation of the Contiguity Model.

	Outputs				
Inputs	A	B	C	D	E
ab	1.00	1.00	0.33		
bc	0.50	1.00	1.00	0.33	
cd		0.33	1.00	1.00	0.50
de			0.33	1.00	1.00

Table 3–7: What the Contiguity Units Compute.

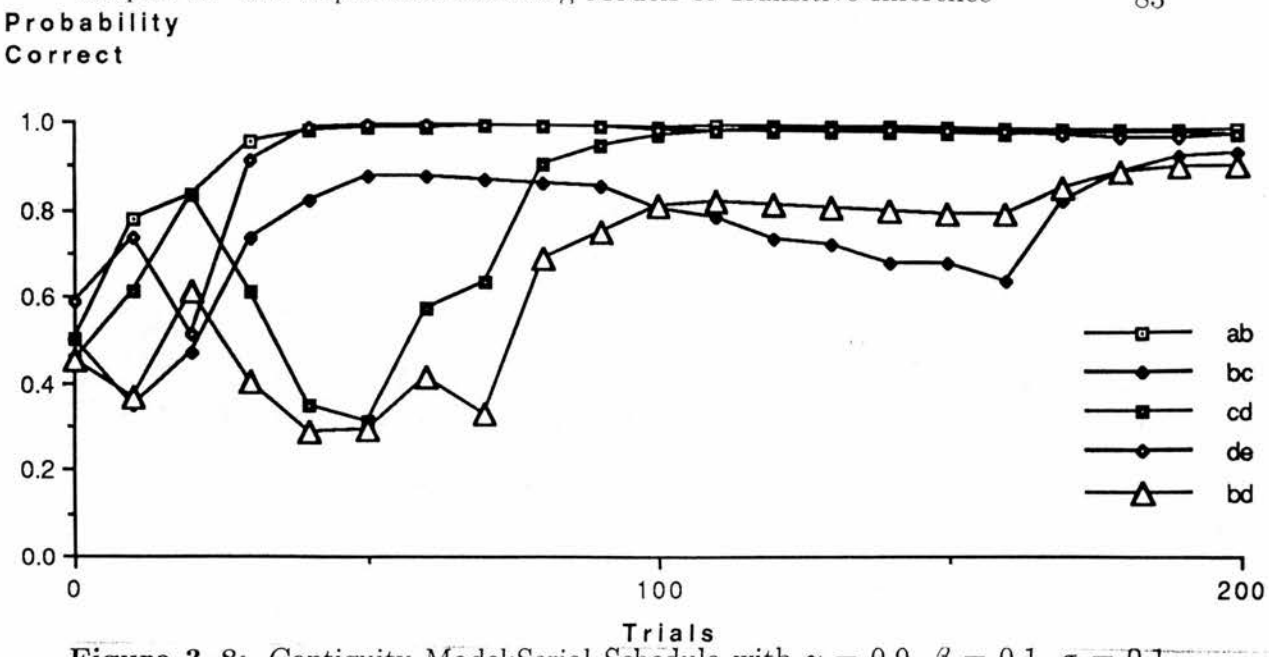


Figure 3–8: Contiguity Model:Serial Schedule with $\gamma = 0.9$, $\beta = 0.1$, $\sigma = 0.1$ and $\rho = 0.1$

The outputs of the contiguity units are shown in table 3–7. It shows that the outputs are not exactly what I planned, but they are sufficient. To get the exact outputs would require a much more complicated error function and learning rules, but I am quite happy to show that much the same effects can be produced by simple rules and representational constraints.

3.7.1 Performance of the Contiguity Model

Figures 3–8 and 3–9 show the performance of the **Contiguity Model** with the two different schedules. We still get a strong transitive bias and there is still no difference between the different training schedules. As a strong transitive bias appears quite quickly with this model, the learning rate for the evaluation function and the contiguity units was kept small so that it is easier to see the path to expertise in the graphs.

Notice from the graphs that the *bd* comparison is the last to reach a high level of performance. This is due to the serial position effect whereby the *bc* and the *cd* pairs take longer to learn than the premises involving end-points.

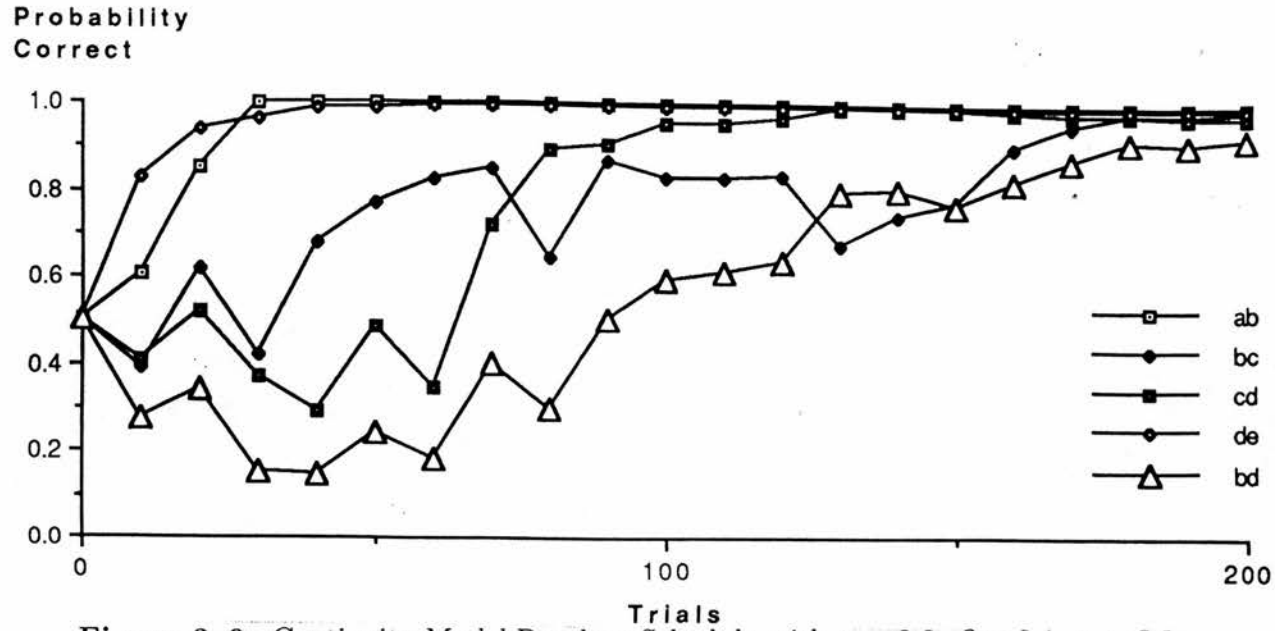


Figure 3–9: Contiguity Model:Random Schedule with $\gamma = 0.9$, $\beta = 0.1$, $\sigma = 0.1$ and $\rho = 0.1$

training	Training and Test pairs									
schedule	ab	bc	cd	de	ac	bd	ce	ad	be	ae
serial	0.99	0.95	0.99	0.99	0.99	0.91	0.98	0.98	0.94	0.99
random	0.99	0.99	0.98	0.99	0.99	0.93	0.98	0.99	0.97	0.99

Table 3–8: Final probabilities from the Contiguity Model $\gamma = 0.9, \beta = 0.1, \rho = 0.1, \sigma = 0.1$.

Table 3–8 below shows the final probabilities of the the test and training pairs from these two runs.

3.7.2 Analysing the Policy

Table 3–9 shows the reinforcements to the weights connecting the contiguity units to the actions.

The general activation from these units will provide a bias in a similar way to the bias units, the mapping units and the memory units. However, we also find that the bD weight will be negative whilst the dB weight will be positive, so the transitive bias is now directly represented rather than being the result of an emergent ranking.

Contiguity Outputs					Actions
a	b	c	d	e	
+	+	+			A
+−	+−	+−	+		B
−	+−	+−	+−	+	C
	−	+−	+−	+−	D
		−	−	−	E

Table 3–9: Grid showing Direction of Reinforcement to Weights between the Contiguity Units and the Actions.

Choice	B	C	D
Serial Schedule	0.63	0.37	0.00
Random Schedule	0.83	0.14	0.03

Table 3–10: Performance on the BCD triad for the two runs shown above

Triads

Furthermore, the performance on the triad *bcd* worse than the test pair *bd*, however the performance on the triad will not repair itself without differential reinforcement.

Table 3–10 shows the performance of the **Contiguity Model** on the *bcd* triad. It shows how the performance has dropped below that of the training pairs and the *bd* test pairs to a similar extent to the monkeys. The differences between the two schedules are not important. The exact performance on the triad varies considerably from run to run, but the drop in performance is a consistent feature.

Retraining on the Triads

The **Contiguity Model** does not repair its performance on the triads without discriminative feedback. However, it is the first model of transitive inference that does change without discriminatory feedback. The contiguity units learn stimulus-

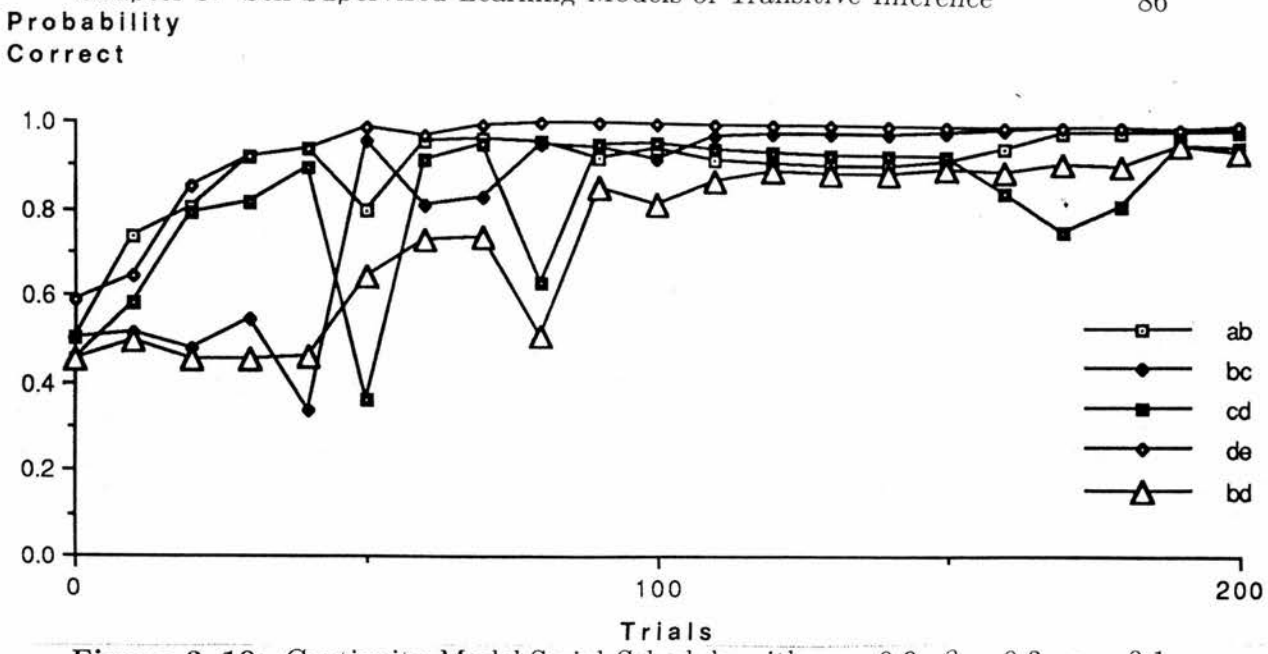


Figure 3–10: Contiguity Model:Serial Schedule with $\gamma = 0.9$, $\beta = 0.3$, $\sigma = 0.1$ and $\rho = 0.1$

stimulus relationships without any feedback whatsoever. When the triads are presented in the dense training phase, these stimulus-stimulus relationships change, and the outputs of the contiguity units will change accordingly. These outputs will affect both the evaluations produced by the critic, and the actual choices made. The reason that the model does not spontaneously repair is almost totally due to the fact that the contiguity units generalise almost perfectly from the binary pairs to the triads so very little change takes place. If the units were instantiated differently such that during the dense triadic phase, the overall activations from the contiguity units increased, then the degree of transitivity would increase. Whether this is the cause of the repair is an open question. Another more interesting possibility is discussed later.

Parameter Values and Training Schedules

It is also possible to generate the observed difference between the random schedule and the serial schedule, but only by pushing the parameters to the limit. Figures 3–10 and 3–11 show the difference between training schedules.

The reason for the failure of the random schedule is simply that a particular pair of stimuli may have no effect on learning for up to seven trials, whereas in

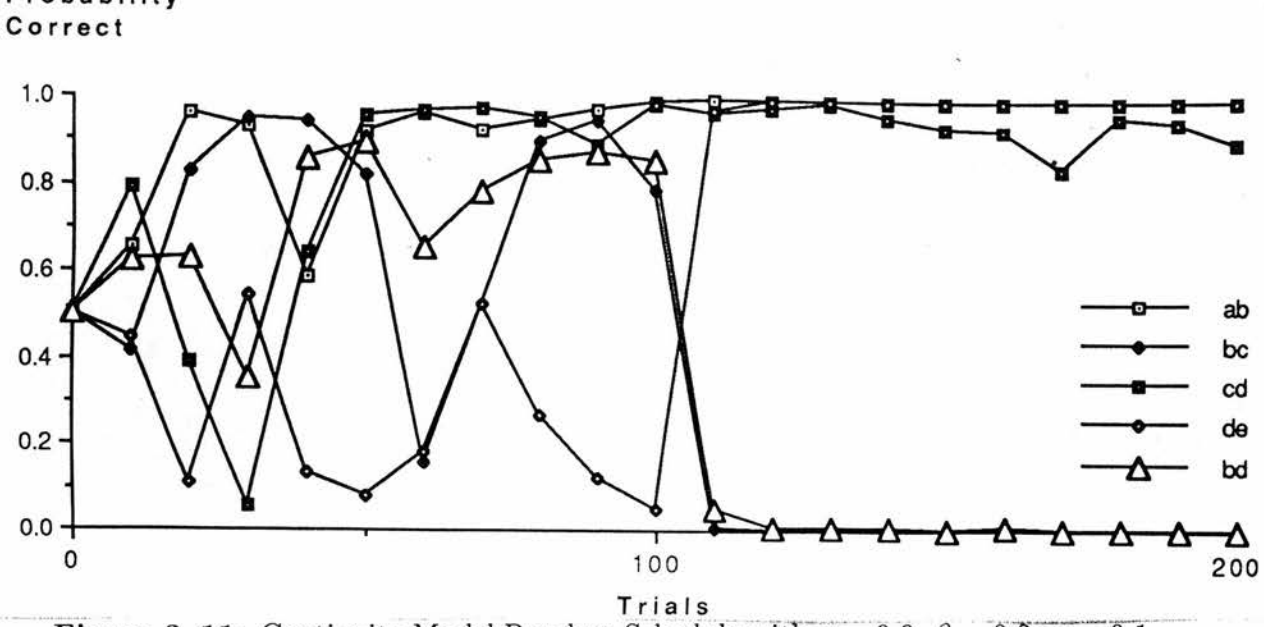


Figure 3–11: Contiguity Model:Random Schedule with $\gamma = 0.9$, $\beta = 0.3$, $\sigma = 0.1$ and $\rho = 0.1$

the serial schedule a particular pair will occur exactly every four trials. There is no reason why the parameters need to be pushed to such a limit when the model works perfectly well over such a large proportion of the parameter space. This robustness is more useful and interesting than the ability to fit one extra experimental effect. This robustness to the parameter values cannot be over-stressed – without pushing the values to unreasonable levels the model will always end up, given enough training, with the same performance.

3.7.3 Conclusions from the Contiguity Model

The **Contiguity Model** is much less parsimonious than the other models. It has more units, more connections and specialist learning rules. It does fit the data better however. It is the only model presented here that shows a strong transitive bias AND a decrement in performance on the *bcd* triad. As I have mentioned before, though, I do not regard fitness and parsimony to be the most important qualities in a model – rather, the insights gained and how the model fits into wider theoretical considerations, are more important.

In the **Contiguity Model**, relationships in the world are modelled and the resulting representation is then usefully applied to improving the performance of

the agent. To claim that the model constructs a representation which is then used meaningfully, might scare Behaviourists, but it has been achieved in such a simple and transparent way that no-one could object to is plausibility as a candidate for a cognitive mechanism. The big question now is whether this investment in complexity can justify itself in improved performance. Obviously, we cannot know the true cost of the investment but we can see if the contiguity units actually hinder performance in any way and if not, then to what extent is positively helps.

3.8 Comparing the Different Models

For the comparisons between the models the parameters were set the same at the following values: $\gamma = 0.0$, $\beta = 0.2$ and $\rho = 0.1$. The value of the learning rate for the contiguity units was set at 0.5.

3.8.1 Which Model Learns Fastest?

Table 3–11 shows how fast the different models learn the four training pairs in the five-term transitivity task. The number of trials required can vary quite a lot; for example, the **Basic Model** learned the premises in less than 160 trials on one occasion but took more than 300 on another. The results are based on the average of five runs of each model. The criterion used was that each premise must be learned to a level where the probability of choosing correctly was greater than or equal to 90%.

Increasing the bias helps to learn a serial relation. The contiguity units also allow the relation to be learned faster. The extent to which the improvement in the **Contiguity Model** is due merely to a bias or whether the particular patterns of activation help is difficult to determine. What we can say though is that the **Contiguity Model** learns the premises in almost 30% fewer trials than the **Basic Model** and throws in a strong transitive bias for free.

Model	Trials
Basic	222
Bias 1	192
Bias 2	184
Bias 3	162
Contiguity	144

Table 3–11: Number of Trials Required to Learn the Premises to Criterion on a Serial Relation

Model	Trials
Basic	344
Bias 1	342
Bias 2	374
Bias 3	370
Contiguity	268

Table 3–12: Number of Trials Required to Learn the Premises of a Circular Relation.

3.8.2 Learning a Circular Relation

All the models can learn a circular set of premises. This indicates that a serial representation is a default and not built in. In other words, unlike the **Value Transfer Model** or Trabasso's **Serial Representation Device**, transitivity is not enforced making it impossible to learn intransitive relations, but transitivity is imposed where it does not interfere with the relation being learned. Table 3–12 shows the performance of the different models in learning a circular set of premises.

We see from the table that adding a bias no longer helps the learning of the premises. If anything it hinders it. The contiguity units, on the other hand, speed up the learning of the premises on a circular relation as well as a serial one.

The **Basic Model** and the **Bias Model** cannot possible generalise appropriately to all the remote pairs in a circular relation. For complete generalisation to the test pairs with distance 2, the following inequalities would have to hold:

	Training and Test Pairs									
	ab	bc	cd	de	ea	ac	bd	ce	da	eb
probability	0.98	0.99	0.97	0.99	0.97	0.83	0.78	0.75	0.74	0.64

Table 3–13: Final Probabilities for Contiguity Model on a Circular Relation.

$$aA > cC$$
$$bB > dD$$
$$cC > eE$$
$$dD > aA$$
$$eE > bB$$

The inequalities are inconsistent; i.e. they cannot all hold. Of course, the inferences could be trained into the network without any trouble, given differential reinforcement. The same is not true for the **Contiguity Model**. Table 3–13 shows how the **Contiguity Model** can generalise usefully on the circular relation, when the premises are learned from scratch.

Not all the generalisations are particularly strong but they are all in the appropriate direction: the direction according to the shortest distance around the circle. Looking at the part of the policy that connects the contiguity units to the actions Shows why. The directions of reinforcement are shown in Table 3–14.

Each colour supports the colour two places above it in the table and undermines support for two places below it. This is the first conclusive evidence that the **Contiguity Model** is not only just a strong **Bias Model** but something extra as well. The **Contiguity Model** can generalise differently depending on what the relation presented to it is. It also learns both relations faster. The **Basic Model** does not generalise strongly at all, and the **Bias Model** can only generalise to a serial relation.

Contiguity Outputs					Actions
a	b	c	d	e	
+−	+−	+	−	+−	A
+−	+−	+−	+	−	B
−	+−	+−	+−	+	C
+	−	+−	+−	+−	D
+−	+	−	+−	+−	E

Table 3–14: Grid showing Direction of Reinforcement to Particular Weights in the Policy for the Contiguity Model Learning a Circular Relation.

Model	Offset	Change
Basic	20	202
Bias 1	224	260
Bias 2	580	130
Bias 3	∞	
Contiguity	164	90

Table 3–15: The Defeasibility of Learning a Serial Relation for Different Models.

3.8.3 Is a Transitive Bias Defeasible?

Table 3–15 shows how the models differ with respect to the ability to learn a circular relation having already learned a series. All the models were first trained on a serial relation for 500 trials. Then the fifth training pair $ae \rightarrow E$ was included. The offset is the number of trials after the initial 500, before the performance on the ae premise drops below 90% choosing A. So the offset gives a measure of how resistant to change the default generalisations are. The change column gives the number of trials required after the offset until all five premises are at 90% correct.

When the bias reaches 3 the serial relation is no longer defeasible. None of the models generalised appropriately to the remote pairs on the circular relation. Whereas, the **Basic Model** and the **Contiguity Model** learned the fifth premise whilst keeping the performance of the other four at a level above 90%, the **Bias**

Models sometimes lost the ability to keep the existing premises at the criterion level. This greatly increased the time for the change to occur. The offset for the **Basic Model** is very small, which is a result of there being no strong automatic generalisations. In both the **Bias Model** and the **Contiguity Model** the offset is considerable. However, the **Contiguity Model** almost makes up for its resistance to change by changing much faster once it starts moving, so the time taken for the **Contiguity Model** to reach the new criterion ends up only marginally longer than the **Basic Model**.

This resistance to change is not necessarily always a bad thing. If a system devotes time and resources towards capturing some perceived order in the world then it is sometimes premature to abandon these assumptions at the first sign of contradictory evidence. Once the evidence mounts up, then of course, it is better to retreat and reorganise as quickly as possible.

3.9 General Discussion

3.9.1 Summary of Results

The **Basic Model** showed that this modelling framework could capture the learning of the premises, the serial position effect, and the drop in performance on the triads without having a strong transitive bias. There is an exceedingly weak stochastic transitive bias and this can be amplified very easily by elaborating the representation with anything at all. The **Bias Model** showed perhaps the simplest possible means of amplification. An order of preferences for touching particular colours emerges over the weights in the policy. As this order gets stronger, (by adding more bias units), the transitive bias gets stronger and the transfer to the triads improves. The subjects on the other hand showed a strong transitive bias without good initial transfer to the triads.

The **Contiguity Model** had elaboratory units that learnt stimulus-stimulus co-occurrence relationships. As the premises present contiguous stimuli they learn

contiguity relationships. This model showed a strong transitive bias *and* a drop in performance on the triads – the only model other than the **Stack Model** to do so. All the models could learn a circular set of premises when presented from scratch but only the **Contiguity Model** could generalise appropriately. The **Contiguity Model** is the only model in existence that can produce a strong transitive bias which is also defeasible.

3.9.2 Self-Reparation on the Triads

It is easy to get a model to perform perfectly on the triads; a strong **Bias Model** is just one example. It is a lot harder to get a model to show incomplete transfer to the triads when performance on the *bd* pair is still high. These models had the potential to capture the incomplete transfer because the support from one colour for choosing another colour was a major part of the policy; the loss of performance on the *bcd* triad is entirely due to the fact that d strongly supports C. The self-reparation, therefore must be due to the fact that this support is either directly reduced or overtaken by the support of colours for choosing themselves in a strict rank.

The difficult question is what causes the change?

There are only two possible candidates for the cause of this change:

1. the increased number of stimuli.

2. the lack of discriminative feedback.

Although, a logical model and a serial representational device model predict that increasing the number of stimuli should help performance, once we consider the task in terms of making choices then obviously the greater the number of options, the harder the choice becomes. This, it would seem, is true despite the fact that the consequences of the different options are the same. But there is of course the possibility that the liberalisation of the feedback allows the application of some extra self-regulatory mechanism, because the subject no longer has to

worry about external constraints and so can concentrate on internal ones. The problem the triads present subjects with is a problem of scope: subjects have the ability to deal with binary comparisons very easily, but once the numbers increase the problem leaves the scope of the decision-making mechanisms. This suggests that the extra self-regulatory mechanisms might be dedicated to bringing within scope more difficult problems. The transitive inference task cannot shed light on this hypothesis however.

3.9.3 Training Schedules

The difference in training schedules was an experimental effect I was confident in capturing with this framework, but I was wrong. With an evaluator designed for evaluating sequences of actions and elaboratory units such as mapping units and memory units which produce different outputs depending on the training schedule, the models still managed to learn the premises when presented in the random schedule. This shows how robust the model is to adding extra elaboratory units. The **Basic Model** could learn the premises with the random schedule without any problem and when extra units are added to the model they do not take this ability away except when the parameters are set to extreme levels. This is an extremely useful property to have in a complex system. It means that we do not have to worry about losing what we already have when we add further units. This extendibility is not assured, however; with the strong bias units, the agent lost the ability to assume transitivity in a defeasible way. Nobody, to my knowledge, has yet come up with a way of incorporating this sort of extendibility as an absolute property, and it looks as if complex systems, both natural and artificial will always have to cope with some limited alternative.

3.9.4 Development

Consider the **Bias Model** with three bias units on the defeasibility task. The model could learn a circular relation from scratch but could not learn it if it was first trained on the subset of premises defining the serial relation. In developmental

terms, this is an example of **vertical decalage**: a developing system has the ability to solve some task and then loses it. Decalage has traditionally been seen as a severe problem for developmental theories, but by understanding development in terms of self-organising complex systems, it is no longer a problem – just a fact of life. An agent is constantly struggling to improve the effectiveness of its behaviour in a complex world, when it has limited resources at its disposal and even more limited information as to what future challenges it will face. It is not surprising therefore that choices are made which are not entirely beneficial. The goal of the agent must be to make choices that, *on the whole*, improve its lot; even if this involves sacrificing some abilities in order to gain others.

The resistance of the other **Bias Models** with fewer bias units and the **Contiguity Model**, to learning a circular relationship having already learnt a serial one, might be interpreted as a case of U-shaped development. The developing system can solve a task, then loses the ability to solve the task and then regains it. In the case of the **Bias Model** there is no important difference in the policy between the success on the circular relation from scratch, and success on the circular relation after learning a serial one, but with the **Contiguity Model** there is the important difference that the agent generalises differently to the remote pairs. Thus, we might interpret the **Contiguity Model**'s performance in terms of an agent reorganizing its knowledge structure in a new way.

These interpretations of the models' performance in terms of well-known developmental effects are not meant to be taken too seriously. The aim is to show that these effects can be accommodated within the interpretation of development as the self-organisation of a complex decision-making agent.

3.9.5 Specific Generalisation Devices

The elaboratory units, whether they are bias, mapping, memory, or contiguity units, can be seen as specific generalisation devices. Some are more specific than others: the bias units can only generalise appropriately on transitive relations whereas the contiguity units can generalise appropriately, at least, on transitive

and circular relations. I argued in chapter 1, that generality was more likely to evolve through this sort of specific generalisation than through a powerful all-purpose problem-solving mechanism. The contiguity units make the model more complex. I have shown, as far as is possible, that this investment in complexity produces benefits that are worthwhile, and therefore agents with this specific generalisation device might be selected for.

In these models, I have only elaborated with one type of unit at a time. Elaborating the representation seems quite robust in that it is difficult to lower the performance through elaboration. It is interesting to note in this respect, that linearly inseparable problems can be made separable through elaboration of the representation but elaboration can never make a separable problem inseparable.

Nevertheless, it seems inevitable that as the number of specific generalisation devices increases the problem of interference will emerge. To counter this, some extra resource management process would need to be introduced.

3.9.6 Conclusions

The ability to make transitive inferences is primitive, but that does not mean that it should not be modelled in a "Piagetian" way. A transitive bias cannot be imposed indiscriminately without taking away the ability to cope with intransitive relations. These models show how a strong transitive bias can be imposed as a defeasible default assumption allowing the agent to learn transitive relations more easily whilst remaining flexible. The simplicity of these models is consistent with the primitive ontological status of the ability, and they also show similar scope to the subjects, as with the drop of performance on the triads.

I have claimed that these models are Piagetian in nature. Self-supervised learning is a type of self-regulation, and the learning is driven through the agent interacting with the task, but if we regard the models as developing knowledge of a relationship between certain items, it is only 'development' on a microscopic scale. It remains to be seen whether this framework of self-supervised learning and

elaboration with specific generalisation devices can be extended to capture larger developmental changes.

Chapter 4

From Transitive Inference to Exhaustive Search via Seriation

4.1 The New Ontology

Piaget was interested in ontology and ontogenesis not just to give a timetable for cognitive development but to inform the crucial questions of cognitive growth. Two major questions are: what advantages does cognitive growth convey and what is the driving force behind it? Until it is known which abilities a cognitive agent starts off with, what then develops from this starting point and in what order, it is impossibly difficult to answer these questions.

Piaget believed that the ability to make transitive inferences relied on high level ordinal abilities that first showed themselves in the ability to seriate. The conventional seriation task devised by Piaget, is one on the most robust indicators of cognitive growth, (Piaget and Inhelder, 1964). The child is presented with ten rods of differing sizes and asked to "make a staircase", or copy a monotonic series placed in front of them. Despite considerable effort to get younger children to succeed at the classical seriation task through various types of intervention and teaching, (e.g. Neapolitan, 1991), it has not proved particularly successful. It is therefore inconceivable that the ability to seriate is a necessary precursor of the ability to make transitive inferences. But transitive inferences rely on a primitive ordering ability. So how can it be that children who can make transitive inferences over a finite set of specific objects, and who do so through some implicit ordering of those items, cannot then seriate those items in a row?

4.1.1 Ordering Abilities and Scope

A possible answer to this conundrum is suggested by the drop in performance on the triadic transfer tests of the transitive inference task. The problem is a problem of scope. The primitive ordering ability that underlies performance on the transitive inference task works well for the binary comparisons required by that task, but fails once the number of items that must be compared increases.

If Piaget had not invented the seriation task for his own reasons and his own ontology, it would have had to have been invented anyway. The triadic post-tests in the transitive inference task, were designed to elucidate the mechanisms underlying transitive choice, but instead they elicited the limited scope. Seriation is a completely transparent ordering task for which the number of items to be seriated can be systematically increased to determine the scope of the underlying ability.

So, what is it that develops that allows older children to successfully seriate large numbers of different objects in order, and what is the driving force behind the development? Whatever it is that develops and whatever the driving force, the advantage it conveys may well be that it allows the cognitive agent to overcome the scope problem. In order to answer these questions it is necessary to understand what is the basis of success on the classical seriation task.

4.2 Understanding the Basis of Seriation

Piaget considered success on the ten-term seriation task as a result of the 'reversibility' of thought. On this view, seriation success requires that the subjects place each item into a series by conceptually coordinating each items relationship with all the larger as well as the smaller items in the set, creating an intermediate value for each item in turn.

Although Piaget's view is consistent with the data from the the classical seriation task, it is by no means the only possible explanation. One problem in

determining the explanatory adequacy of Piaget's theory, is that success on the classical seriation task does not necessarily rely on just one skill, and therefore it is impossible to interpret the reasons behind failure for a particular subject. Failure may well be the result of an inability to calculate the correct ordinal positions, but it could equally be due to ergonomic difficulties, limitations in the size of production a subject can cope with, or just a plain misunderstanding as to what is required, to name but a few. Without knowing the reasons for failure, we cannot be sure about the basis of success.

4.2.1 Decomposing Seriation Skills

To increase our understanding of the basis of seriation skills McGonigle(1987), designed a task to test the ordinal size comprehension of subjects independently from the serial production elements of the classical seriation task. More recently, Chalmers and McGonigle(in press), decomposed the classical seriation task into four subtasks using learning procedures and touchscreens. Learning procedures give a richer and more easily interpretable data set, from which to determine what an individual subject can or cannot do, whilst the use of touchscreens avoids the worst of the ergonomic problems.

The four tasks were:

Size Seriation The subjects had to learn to touch stimuli of different sizes in a particular order.

monotonic The correct sequence was either to touch the biggest first then the next biggest and so on down to the smallest, or the other way round. This task is the most similar to the classical seriation task.

nonmonotonic The correct sequence was one of 32514 and 43125, where 1 refers to the smallest, 2 to the next smallest and so on upto 5 being the biggest. Monotonic series seem to be so ingrained in our minds that nobody had considered nonmonotonic seriation as a valid type of seriation before this.

Colour Seriation The children had to seriate an arbitrary colour sequence. The items to seriate are more easily discriminable than in the size seriation tasks, but there is no perceptual information which the subject can pick up to guide their next choice. Thus, this task might be considered as testing seriation skills in their purist form.

Matching-to-Sample Two size series were presented on the screen but the actual sizes of the corresponding items, (according to ordinal position), were different. One item in the sample series would be flashing and the subject had to touch the item in the other series with the same ordinal position. This task does require sophisticated ordinal skills.

A schematic representation[1] of these tasks is shown in Figure 4–1.

Figure 4–2 shows the relative performance on the four tasks for two different age groups. The first thing to say about the results is that the seven year olds did better than the five year olds on all the tasks. Five year olds found monotonic size seriation and colour seriation relatively easy, and the nonmonotonic seriation and matching-to-sample very hard. This suggests that seriation does not rely on high-level ordinal skills, in fact some five year olds found colour seriation easier than monotonic size seriation where the monotonicity makes the ordinal position map on to the sequence position directly. The most significant improvement between the five and seven year olds was in the matching-to-sample task indicating that seven year olds do have a much better understanding of ordinal positions, but the fact that they still found the nonmonotonic size seriation task so difficult shows that even with the ability to calculate ordinal positions, it is not used extensively in seriation.

Perhaps more important than the differences between the two age groups is the continuity. At both ages, colour seriation and monotonic size seriation are relatively easy and nonmonotonic size seriation is difficult. So if the improvement

---

[1]reproduced with permission from Chalmers and McGonigle(in press)

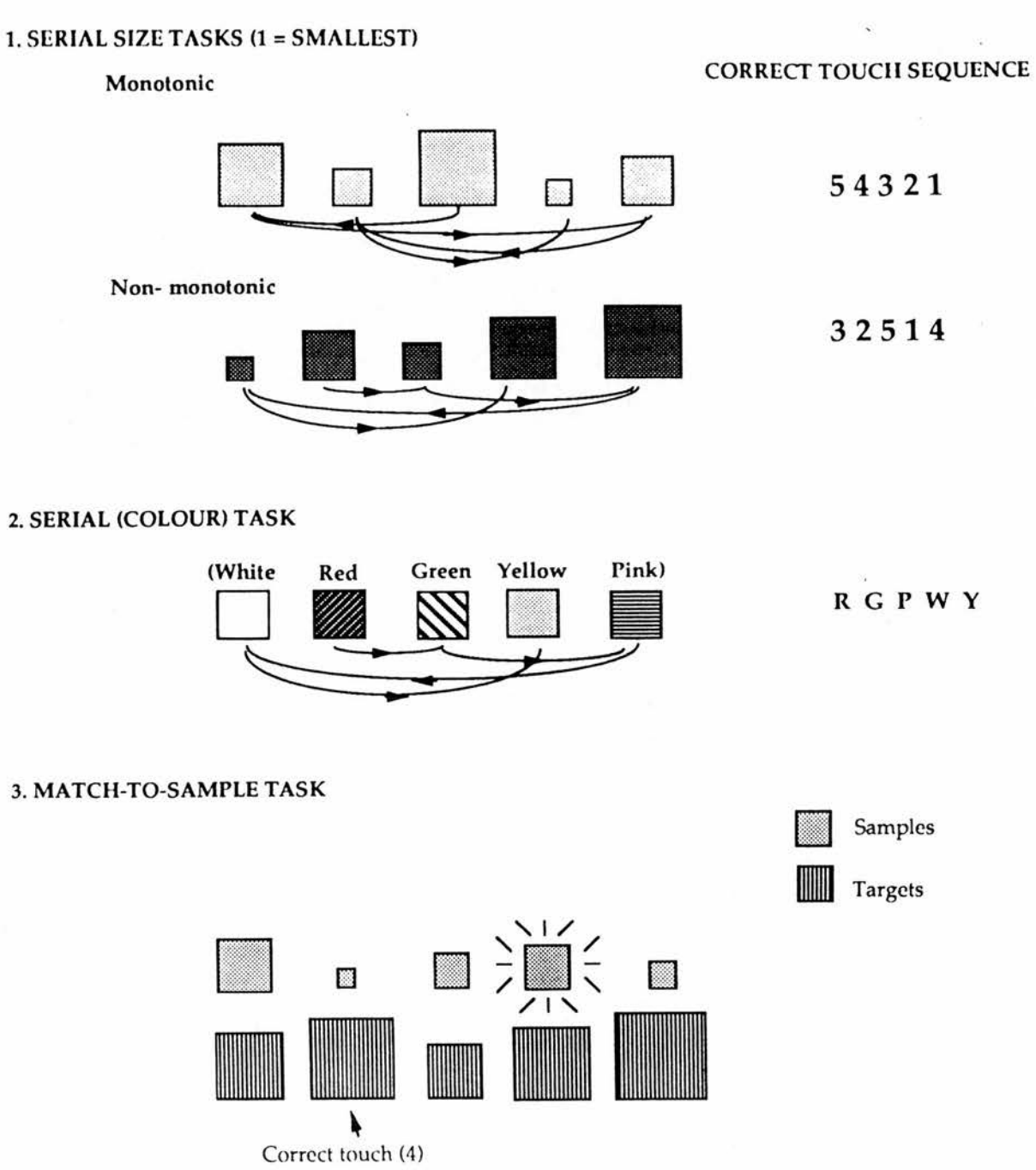


Figure 4–1: Schematic Representation of the Four Tasks Used to Decompose Seriation Skills

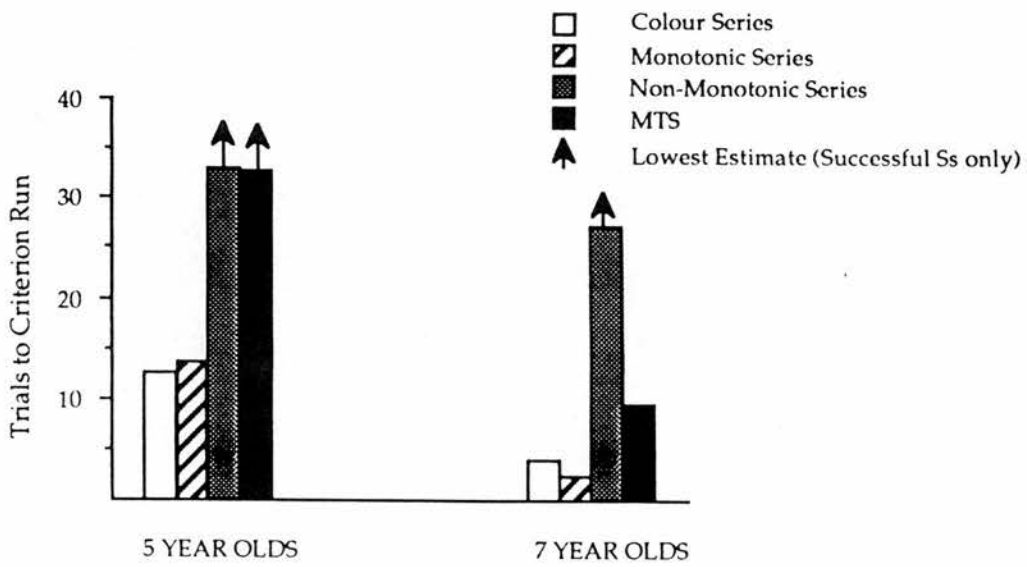


Figure 4–2: Performance on the Four Tasks for Two Different Age Groups

on the easier tasks is not the result of higher-level ordinal abilities then what is the cause? Chalmers and McGonigle suggest that the improvement is the result of specific strategies that reduce the 'cognitive strain' on the subjects. Thus, with the monotonic series, the good subjects pick up the redundancy afforded by the stimuli, and then search for the next item to be touched according to the direction of the change and the interval of difference between adjacent items. Using this strategy the subject no longer has to remember exactly what has been touched so far and what has not. All they need to know is the size of the last stimulus touched and with this they can direct their search for the next item. Of course, this strategy cannot be applied to the nonmonotonic case, which would explain why it remains so difficult. Furthermore, Chalmers and McGonigle found that the best predictor of errors in the nonmonotonic task was the principle of choosing an adjacent size. This is evidence of a data-reducing strategy wrongly applied.

The pattern of the results suggests that children are developing strategies that use the redundancy afforded in the monotonic task in order to reduce the cognitive strain imposed by working memory.

This is consistent with what Wright(1985) found with human adult subjects. Their problem solving expertise improved with experience and was often the result of data-reducing strategies. After considerable practice on three-term transitive inference problems, the subjects devised mneumonics such as eliminating all unnecessary information, and therefore reducing the load on memory. For example, they might retain only "J and B" from the statement "Jennifer is shorter than Brosnan". These strategies are not suggested to the subjects, rather the subjects spontaneously devise them.

4.3 A New Characterisation of Development

4.3.1 Scope and Data-Reducing Strategies

The picture of cognitive development emerging from this programme is changing dramatically. We do not develop new skills so much as develop strategies that reduce complex tasks to simpler ones we already know how to do, thus extending the scope of those basic abilities leading to a richer behavioural repertoire. The difficulty of producing a data-reducing strategy will depend mostly on the specific details of the task. This suggests that development will not progress in a smooth, monotonicly increasing graph but in a rather haphazzard way leading to peculiar discrepencies such as the ability to seriate a monotonic size series without the ability to seriate a nonmonotonic one.

To say that we have a finite set of basic abilities of limited scope is slightly misleading. The basic choice ordering ability that underlies transitive inference, and the basic search ability that underlies successful monotonic size seriation, are limited in scope in terms of efficiency. In terms of potential applicability they are both extremely general, and this is crucial. It has been stressed throughout this thesis that a developing system must be competent and robust at all stages of its development. If an agent finds itself in a difficult situation, then the most important aspect is not whether the agent can solve the particular problem at hand in an optimal or even an efficient manner, but whether it can solve the problem at all, and so extricate itself. The agent cannot just sit tight and wait for some new ability to develop, it must start developing strategies immediately.

4.3.2 Applying Constraints and Resource Management

If the purpose of development is to increase the scope of our limited cognitive abilities, then the driving force is unlikely to be some abstract concept like truth preservation. Of course the need to solve new tasks can be a driving force in itself,

but the seriation work, intriguingly suggests that resource management may play a crucial role. Subjects solve the monotonic size seriation task by applying a data-reducing strategy that reduces the task to a simple search task, and thus reduces the strain on working memory. The question is whether this reduction in 'cognitive strain' is simply a pleasant side-effect or whether it plays a causal role in the development of the strategy to begin with. The latter certainly sounds more plausible.

Returning to the transitive inference paradigm, we now have an explanation for the monkeys' ability to self-repair their performance on the triads. The triads illustrated the scope problem, and the self-regulation in the absence of discriminatory feedback can be understood in terms of some resource management process that increases the scope of the simple choice ordering device.

The important questions now are:

1. why should resource management play such an active role in the development of data-reducing strategies?

2. how does this process work?

3. how could we investigate this possibility empirically and through modelling?

If we just passively respond to constraints placed on us by the environment and our limited cognitive resources, and this adaptation leads to the progressive constraining of our behaviour, then sooner rather than later this adaptive process is going to grind to a halt. We may have become more *adapted* but only at the cost of becoming less *adaptive*. This may well be an accurate depiction of development in canines, ("you can't teach an old dog new tricks"), but it does not reflect what we see in humans. People can change their view of the world, solve new problems, and adapt to new environments throughout their lifespan, barring senility or other mental impairments. So how do we do this? Obviously, by using intelligent resource management processes, we can become constrained in the most efficient way possible.

So a data-reducing strategy might develop through deliberate constraining of the agent's behaviour, which may limit its potential scope, but increases its actual scope by increasing the efficiency of its basic abilities in specific domains. The specific problems in a specific domain will have a great impact on how the agent becomes constrained and so will active resource management processes within the agent itself.

The ability to generalise to other tasks using a strategy learned on a specific task will depend on the relationship between the task and the sorts of tasks that agent has evolved to tackle. If the strategy has been developed to solve a task with no ecological validity, then it is likely that there will be no useful generalisation to be derived from it. If, on the other hand, the strategy was developed to solve an ecologically valid task, then the chances of being able to generalise to other ecologically valid tasks is much greater. This whole research programme has moved more and more to using ecologically valid tasks. For example, the transitive inference task was extended to give triadic tests because binary decision making was determined to be artificially constraining. Now it is necessary to adapt the seriation task to allow self-regulation. Self-regulation has been the main focus of the modelling, and for this reason the seriation tasks, described so far, where subjects are explicitly trained to produce a particular sequence of responses is not modelled in this thesis.

4.4 Designing a New Task

To understand the development of data-reducing strategies it is necessary to develop new experimental paradigms. The tasks should have liberal reinforcement regimes to allow the subject sufficient freedom to regulate their own behaviour rather than them being 'shaped'. Similarly, the task should not have a 'correct' solution but a wide distribution of 'good' solutions, so we can observe which possible solutions the subjects find, thus indicating how the subjects deliberately constrain themselves. The task should be able to be presented again and again so

task experience becomes one of the main variables. And finally, it must be possible to systematically alter the difficulty of the task in order to stretch the subjects and force them to self-regulate.

McGonigle (McGonigle, De Lillo, and Dickinson, 1992), has devised such a task: the exhaustive search task. To complete a trial in the exhaustive search task every stimulus in a set must be searched. However, it is left upto the subject to decide which order they should be searched in. The subject is free to make as many touches as they like to the same stimulus although additional touches are redundant. The subject is rewarded at the end of every trial no matter the efficiency of the search. By giving the same task over and over again, the subject is allowed to self-regulate their performance such that they make fewer redundant touches, and thus their searches become more efficient. And the number of stimuli to be searched can be systematically increased to make the task harder. The main hypothesis is that subjects will constrain their searches such that they consider a decreasing proportion of all the possible minimal search paths, and that these search strategies will lead to a greater efficiency of performance.

4.5 The Modelling Strategy

Modelling must play a crucial role in this research programme. The mechanisms underlying self-regulation and resource management are very complicated, and without the ability to simulate our theories we will never be able to tell whether they sufficiently account for the behaviour we observe. However, there are severe problems in modelling this sort of process. Most severe is the problem that we do not know what cognitive resources are available to the subjects. Trying to model resource management without knowing what resources are being managed is a mug's game. Therefore, it is first necessary to determine what resources are required to solve the task. Having determined a minimal set of resources that could potentially succeed, we can then investigate how the sorts of constraints that subjects use, can improve the efficiency of their searches. Once we have a

rough idea about the resources and the constraints, it should be possible to infer how resource management might lead to these constraints given these resources.

4.5.1 The Modelling Framework

This modelling strategy is highly exploratory, and we need a modelling framework that allows resources and constraints to be added as easily as possible. We also need a framework which can be naturally applied to search problems. Sutton and Barto's **Temporal Difference Learning Model** fits the bill perfectly on these two counts. The transitive inference modelling showed clearly how easy it was to elaborate the model through the representation. Thus, we can add resources through elaboration. Temporal difference learning was designed to capture sequencial decision-making which can be, very naturally and easily, applied to search problems. Finally, one of the most crucial parts of the experimental paradigm is the minimal feedback, and with the critic modifying the reinforcement in the way that it does, we have a modelling framework with the potential to cope with the task.

4.5.2 The Need to Keep Things Simple

This exploratory modelling methodology is powerful in that it allows the modeller to try a huge number of variations with very little effort. To control this power there are two important constraints that the modeller must apply vigorously at all times. Firstly, the modeller must be honest. It is easy to introduce hacks to solve particular problems and at times this is a sensible thing to do, but the modeller must make it absolutely clear that this is what they are doing. Secondly, because the models can become quite complicated very quickly, it is essential that the modeller understand not only what the models are doing but also how they are doing it. For this reason it must be as easy to analyse the models as possible. If the activation rules are non-linear then it can be extremely difficult to work out the relationships between the input activations, the values of the weights, and the activations produced over the output units. Given this situation,

then if the model is doing what it is supposed to, the modeller still cannot be certain that it is doing it for the right reasons. If on the other hand the model is behaving inappropriately then the modeller has no clues as to what changes need to be made. Making the activation rules linear, (e.g. the output of a unit is simply the sum of the products of its inputs and the corresponding weights), makes the relationships between activations and weights much easier to interpret. Of course, as shown by Minsky and Papert(1969), this means that the models cannot cope with linearly inseparable problems. However, linear inseparability can be overcome through elaborating the inputs of the network with appropriate units and activations, which of course, is part of the modelling strategy anyway.

Chapter 5

Exhaustive Search in a 3x3 Grid: the Task

5.1 Introduction

The decomposition of the seriation task described in the previous chapter suggested that what develops is data-reducing strategies that increase the scope of existing abilities, and that these strategies develop by means of self-regulation driven by resource management. The seriation tasks, however, did not allow the subjects to self-regulate – the subjects were driven through particular serial productions by a rigid reinforcement regime – and therefore these tasks cannot shed light on the self-regulating processes themselves. To understand these we need a new task and for this reason the exhaustive search task was designed. Given that the major failing of the transitive inference models was their inability to capture the subjects ability to self-regulate their performance on the triads without differential reincement it is also hoped that this new paradigm will illuminate exactly what it is that is lacking in these models.

Before modelling can begin, it is necessary first to analyse the task to allow proper evaluation of the subjects' performance, and secondly to characterise the subjects performance to direct the modelling and allow proper evaluation of the models produced.

In this chapter the experiment is described. Random models are produced to provide a baseline to evaluate the subjects' and the models' performance. Some

simple analysis of possible strategies is given which highlights some of the difficulties faced by the subjects. Finally, the initial characterisation of the subjects' performance that directed the modelling at first, is presented.

5.2 The Experimental Procedure

The experimental procedure is that reported by McGonigle et al(1992).

A number of stimuli (monochrome squares) are shown on a touchscreen in random positions in a 3x3 grid. The subject has to touch every stimulus before receiving any feedback. Once all the stimuli have been touched, the subject receives a reward and the screen goes blank for a while before the presentation of the next set of stimuli. 12 children (2.10 yrs -4.7yrs) and six brown capuchin monkeys (*cebus apella*), were subjects. For both species of subject the stimuli were green monochrome squares approximately 40mm across. These were placed on the screen randomly in one of nine positions, (three rows of three). For each touch to a stimulus the latency and the position of the stimulus were recorded. Also a registration signal was given consisting of a short tone, and either a flash of the stimulus touched or the brief presentation of a ring around the stimulus.

The procedures were slightly different for the two species and the differences are detailed below.

5.2.1 Procedure for the Children

The children were shown one stimulus on the screen and encouraged to touch it with an outstretched finger. At no point in the experiment were any other directions given to the subjects. After the registration signal, the screen went blank, and the childrens 'reward' was given, as one touch in this situation is all that is required for an exhaustive search. The childrens 'reward' was to be shown a computer animated carton image. The image showed a figure climing a ladder in order to reach a piece of fruit hanging from a tree. If the subject had made no

redundant touches, the figure would climb one rung closer to the fruit. Six minimal trials were required for the figure to reach the fruit. Once the figure reached the top the subject would be given one extra stimulus to search on the next trials. Subjects were taken through the incremental item number arrays from one upto nine items per trial, dependent on their degree of success and task-motivation as determined by the experimenters.

5.2.2 Procedure for the Monkeys

The monkeys first needed to be shaped to touch a stimulus on the screen. Once this had been done they were given daily sessions of upto fifty trials, all with the same number of stimuli. The subjects responses and speed of responses were recorded. Several statistical measures of the subject's performance were extracted for each session and one, the percentage number of correct responses, was picked to be used as part of the criterion for deciding when a particular monkey should be advanced onto the next stage of the task. A correct response was considered as a touch to a stimulus that has not been touched before, so if on a particular trial a monkey makes five responses when there are only four stimuli then the percentage number of correct responses for that trial is 80%. When a monkey touched a stimulus in the same position twice in a row it was not clear whether it was a deliberate touch or a type of stutter. For the sake of the criterion, then, immediate repeats were ignored. If a subject scored over 75% correct responses over a complete session then they were deemed to have reached criterion. Unless their performance was still dramatically improving they were advanced onto the next stage once they had reached this criterion.

Having reached the criterion of performance on two stimuli the subjects were then given three, and so on upto five stimuli. This constituted **Phase I** of the task. **Phase II** of the task involved giving the subjects different numbers of stimuli in the same session. Five different numbers were given from four stimuli upto nine stimuli. In each session there were five blocks of ten trials with each block having the same number of stimuli. The order of the blocks in each session

changed according to a latin square design. **Phase II**, dubbed the **embedded phase**, lasted for two weeks with five sessions in each week. In the first week the numbers of stimuli used were four to eight inclusive, and in the second week, five to nine. **Phase III** of the experiment consisted of giving the subjects nine stimuli for complete sessions of fifty trials.

5.3 Task Analysis

This experimental paradigm is new and it is necessary to examine the task in some detail both in order to properly evaluate the monkeys performance, and to establish what minimal set of competences an agent would need to solve the task.

5.3.1 Random Searches

The statistics needed to thoroughly analyse the performance of the monkeys, (as well as the models of the monkeys), do not exist. Random searches provide a baseline with which we can compare the monkeys' and the models' performance. Three random search models were produced:

Basic Monte Carlo Model In this model the choice of stimulus to touch next at each point in the trial is produced by a software random number generator, such that each of the stimuli presented in that trial is equally likely to be chosen.

Probability Model This model also makes the assumption that each stimulus presented is equally likely to be chosen at each point in the trial, but it is a probability model. Hence, there is no sampling bias produced by a random number generator. The Monte Carlo Model was produced because this probability model was difficult to work out and took a long time to produce. Even though it was late in arriving it does allow evaluation of the Monte Carlo Model. The mathematics behind this model are shown in the next section.

Monte Carlo Random Walk Subjects showed a tendency to move to an adjacent stimulus if one existed. This model assumes a starting point in the grid and a random walk through the grid, where each move is a horizontal or vertical move to an adjacent grid position. Whenever, the agent is in a grid position where there is a stimulus, then that stimulus is touched.

The Random Probability Model

1

Assuming that each stimulus is chosen at random at each point in the trial, what is the probability that the agent will receive the reward (i.e. have exhausted the set of stimuli) on turn T? The key to working this out is to forget that the trial ends when the set is exhausted and assume that the trial just continues on and on.

$A(N, M)$: The event that in N goes the agent will have touched all M stimuli.

If we know this, then we can work out the probability that it gets the reward on turn T, because this is simply $p(A(T, M)) - p(A(T-1, M))$, the probability that it has touched them all by turn T minus the probability that it has touched them all already by turn $T-1$.

To solve this consider the number of sequences which the agent can make to satisfy $A(N, M)$.

The total number of possible sequences is M^N, but we have to subtract off this the number of sequences which don't involve touching all M stimuli. Any sequence will touch precisely i stimuli for some i in $1, 2, ..., M$. Let us call the number of stimuli a sequence involves (touches) its index. So the index of

a a a

is 1, while the index of

---

[1]The mathematics in this section was done in collaboration with Steve Finch.

a b

is 2.

Now, fix the length of the sequence (N above). From all sequences of length N, we want to find those with index M. This is all sequences of length N less those with an index of $M-1, M-2, \ldots, 3$, 2, or 1, (since every sequence must have some index in $1, \ldots, M$).

But the number of sequences of length N with index $M-1$ is simply the number of sequences satisfying $A(N, M-1)$, and so on. Well, that's almost true. We need to allow for the fact that if we have M stimuli then here are many different sets of $M-1$ stimuli the sequence could have involved. In general if we consider $A(N, m)$, there are $\begin{pmatrix} M \\ m \end{pmatrix}$ sets of stimuli. Since no sequence with index m can be in any two of the sets of m from M stimuli (because ALL of the stimuli have to be included in the sequence), we count no sequence twice in this analysis. So the total number of sequences with index m we have to subtract is $\begin{pmatrix} M \\ m \end{pmatrix} A(N, m)$.

Summing over all the indexes we have to take away from the original number of sequences, we find

$$A(N, M) = M^N - \sum_{m=1}^{M-1} \begin{pmatrix} M \\ m \end{pmatrix} A(N, m).$$

This is the total number of sequences with index M.

The probability of having won by turn N with M stimuli (assuming we carry on playing after we win) is simply

$$\frac{A(N, M)}{M^N} = p(N, M)$$

and after a little bit of maths this boils down to the following recurrence relation:

$$p(N, M) = 0 \text{ for } M > N$$

$$p(N, 1) = 1 \text{ for } N > 0$$

Basic Monte Carlo Model							
Number of stimuli	3	4	5	6	7	8	9
Average trial length	5.6	8.6	11.2	14.3	17.8	21.9	25.7
percentage of correct touches	53.2	46.4	44.8	41.9	39.2	36.6	35.1
percentage of minimum trials	22.0	8.4	4.4	2.4	0.4	0.3	0.1
Probability Model							
Number of stimuli	3	4	5	6	7	8	9
Average trial length	5.5	8.3	11.4	14.7	18.2	21.7	25.5
percentage of correct touches	54.5	48.0	43.8	40.8	38.6	36.8	35.3
percentage of minimum trials	22.2	9.4	3.8	1.5	0.6	0.2	0.1
Monte Carlo Random Walk							
Number of stimuli	3	4	5	6	7	8	9
Average trial length	6.6	10.3	14.7	19.5	23.9	28.2	32.8
percentage of correct touches	45.3	38.8	34.0	30.7	29.3	28.4	27.4
percentage of minimal trials	17.0	6.7	3.2	1.1	0.6	0.3	0.0

Table 5–1: Results of the Random Models

$$p(N,M) = 1 - \sum_{m=1}^{M-1} \binom{M}{m} \frac{m^N}{M^N} p(N,m) \text{ otherwise.}$$

As mentioned before, the probability of winning ON turn T is simply $p(T,M) - p(T-1,M)$.

Performance of the Random Models

For the basic Monte Carlo model 1000 trials were given. For the Monte Carlo Random Walk 900 trials were given, 100 starting from each grid position.

Table 5–1 shows the performance of the random models. The basic Monte Carlo model gives comparable results to the probability model on all of the measures shown, so at least we know that the random number generator is reasonably good.

The average trial length increases steadily as the number of stimuli increases. The percentage of correct touches drops rapidly to begin with, but the rate of decrease slows as the number of stimuli increases.

The percentage of minimal trials drops even more rapidly than the percentage of correct responses.

The Monte Carlo Random Walk does worse on every measure, so adjacency on its own is definitely bad. It seems therefore that subjects' tendency to move to an adjacent stimulus must either be part of some complex strategy for which adjacency is only one part, or adjacent moves have some other positive aspect. One possibility of course is that the monkeys rate their performance on the the total distance moved rather than just the number of touches required; thus a sequence with a higher proportion of adjacent moves would be preferred to a sequence of equal length but with a lower proportion of adjacent moves.

It is important to note that the random models do actually solve the task, i.e. they do eventually touch all the stimuli, even if it can take a long time.

5.3.2 Configuration Effects

With most search tasks it might be expected that the difficulty of the task increased monotonically as the number of items that needed to be searched increased. In this task however this effect is confounded by the number of configurations of stimuli that can be produced by different number of stimuli.

The number of configurations for n stimuli corresponds to the the number of ways of picking n items from 9 when the order of the picked ones is unimportant.

Number of combinations = $\frac{9!}{(9-n)!n!}$

Number of stimuli	1	2	3	4	5	6	7	8	9
Number of configurations	9	36	84	126	126	84	36	9	1

Although there are the same number of configurations for four and for five stimuli, there are a lot of configurations of four which are very similar to each other; i.e. simple transformations of each other. If we assume subjects can use

simple transformations such as translations to effectively reduce the number of configurations then the effective number of configurations will change. The condition with 5 stimuli, will have the highest number of effective configurations by quite a considerable margin.

5.4 Possible Strategies

Consider a *strategy* as something that informs an agent about what is a good thing to do at each choice point during a trial. Given this definition, the statement, "Do not repeat or reiterate!" does not constitute a strategy as it only tells the agent what *not* to do. A complete strategy with this as a component might be: "Move randomly from one grid location to another, and if the location contains a stimulus which has not yet been touched then touch it". This might be dubbed the **Negative Strategy**. If an agent were to follow this strategy it would always produce minimal trials. However, there are a number of other criteria that should be taken into account when considering the efficacy of a particular strategy. Firstly, it is important to take into account the cognitive resources that an agent would need in order to follow the strategy. Secondly, the agent may not measure its performance purely by the number of minimal trials it produces but use some other measure such as the time taken between the beginning of a trial and receiving a peanut. Thirdly, there is no point considering a strategy for an agent if it is impossible for the agent to instantiate that strategy or if the strategy is impossible to learn.

Given these criteria the Negative Strategy does not look so good. It relies very heavily on the agent being able to remember every location where a stimulus has been touched; thus it requires a perfect memory for a potentially huge number of items. Also, if we consider that moving to a location incurs some cost, then the strategy is not even particularly efficient; although the number of touches is minimal the number of moves is not. In its favour, though, the negative strategy is probably relatively easy to learn.

1	2	3
4	5	6
7	8	9

Figure 5–1: Reference Numbers for Grid Locations

Better strategies would rely less on working memory and produce fewer moves. Assuming the agent can differentiate between the different grid locations according to figure 5–1, then a better strategy might be: "If no stimulus has been touched then touch the stimulus at position 1, if the stimulus at position 1 has just been touched then touch the stimulus at position 2, ..." and so on up to position 9. This strategy would be optimal when nine stimuli were presented, both in the number of moves made and touches required. Furthermore, it only requires the agent to remember the last location touched, and so it does not place an unreasonably high burden on working memory. However, what would happen if there was no stimulus in, say, position 2? The strategy would collapse.

The strategy, then, must always be able to suggest a possible action no matter what the configuration or the number of stimuli presented.

5.4.1 Preference Rankings

The strategy above can be adjusted to make it work for all task situations by making the action suggested into a rank of preferred actions. The action chosen is the most preferred action which is applicable. Consider the table of rankings in Table 5–2.

The numbers across the top represent the last position of the nine that was visited and the columns below represent a ranked preference for moving to another position. So, if a subject started by touching the stimulus at position 1, and the subject was using this set of rankings to decide their choices, then the table tells us that the subject would go and touch the stimulus at position 2. If there is no

Position last touched	1	2	3	4	5	6	7	8	9
	2	3	4	5	6	7	8	9	1
	3	4	5	6	7	8	9	1	2
	4	5	6	7	8	9	1	2	3
rank of positions	5	6	7	8	9	1	2	3	4
to be touched next	6	7	8	9	1	2	3	4	5
	7	8	9	1	2	3	4	5	6
	8	9	1	2	3	4	5	6	7
	9	1	2	3	4	5	6	7	8
	1	2	3	4	5	6	7	8	9

Table 5–2: An Optimal Positive Strategy Implemented as a Set of Rankings

stimulus at position 2 then the subject would try the next item in column 1 which is 3. If there is no stimulus at 3 then they would move on down the column etc.

Table 5–2 gives one correct solution, but how many others are there, and what proportion of the inductive space do they take up? Consider the transformations you can apply to the table that preserve the non-looping. When all nine stimuli are present only the top row of the table needs to be considered. There are nine possible starting positions and having chosen one there are eight other positions which would not be an immediate repeat. Having chosen one of these, there are seven alternatives that do not repeat or reiterate. So the number of different non-looping alternatives for the top row is 9!. The total number of alternatives for the top row is 9^9. Therefore, the proportion of correct solutions for nine stimuli is:

$$9!/9^9 = 9.367 \text{ x } 10^{-4}$$

Now consider the case with eight stimuli. The top two rows of the table must be used and trying a few simple manipulations of the second row quickly reveal that it is totally determined by the top row. When considering seven items it is necessary to look at the top three rows and again there is only one possible configuration that maintains the non-looping non-repeating behaviour. This can be continued

down to the case with one stimulus where the the ninth row is totally determined. So to be fully competent with any number of stimuli and starting from any point in the grid an agent still has to find one of 9! solutions, but the number of different possibilities now are $(9!)^9$. Therefore, the proportion of correct solutions becomes:

$$9!/(9!)^9 = 3.326 \text{ x } 10^{-45}$$

Knowing the size of the inductive space is not in itself sufficient to decide the learnability of the solution. It is necessary to take into account a number of other criteria such as the time available to find a solution, the richness of the feedback as to how far you are from the solution and relatedly, the contours of the error space and the information that can be used as to the general direction of the solution. In this task all of these criteria only serve to underline the difficulty of the problem. The subjects only get fifty trials a day so if they want to improve they must do so quickly. The feedback is as impoverished as it possibly could be, whilst still defining the task. And the error space is very unpredictable - an agent may be incredibly close to a solution and receive very little reward, and very far away and receive much more.

Consider just the case for learning to solve the task with nine stimuli. It is only necessary to use the top row of the table. Now imagine that the top ranking choice in column 1 is 3 instead of 2. Let all the other columns remain the same.

position last touched	1	2	3	4	5	6	7	8	9
rank of positions to be touched next	3	3	4	5	6	7	8	9	1

We might consider this state to be only one step away from the solution, but consider its performance. If the agent starts at position 2 then the agent's performance is still optimal but if it starts from anywhere else it will go round in a big loop until either its just comes to expect that it is doomed to receive no more peanuts or it makes one of the other transitions go to position 2. The problem is that there is no way of telling which of the transitions within the loop is the bad one, and if one is changed then there is only a one in eight chance that it will be the correct transition from position 1 to position 2. If it is one of the others then the strategy has moved to a state where it is two steps away from the ideal

solution. Even if the transition from position 1 to position 2 is finally changed the damage to the current policy has already been done; The other transitions are likely to be so weakened that other mistakes start creeping in and the little bit of good has been swamped by a great deal of error which makes it very difficult for the agent to sort the good from the bad.

Even if an agent had fluked coping with nine stimuli, and was then moved on to eight, the errors that it would make on eight could not be categorised into those affecting the eight stimuli task and those affecting the nine stimuli task and so it would unlearn nine as soon as it was given eight.

5.4.2 Summary

Random behaviour is not particularly efficient, especially as the number of stimuli increases. However, although a random search is not particularly good, it is robust – some trials may take a huge number of touches but the set *will* be exhausted eventually.

A negative strategy, involving decisions about what not to do, can lead to efficient behaviour, but relies heavily on working memory. As using working memory increases cognitive strain and is limited anyway, we might expect subjects to try and adopt some positive strategy.

A positive strategy for searching the stimuli, reduces the load on working memory considerably, but unless the subject gets the 'right' positive strategy, there is the possibility of catastrophic looping. Furthermore, with the instantiation of a positive strategy as preference rankings, it is incredibly difficult to acquire the 'right' positive strategy.

So a random strategy is inefficient yet robust, a negative strategy can be efficient but is very expensive, and a positive strategy would be efficient and relative cheap, but is extremely difficult to acquire and the cost of failure is catastrophic. Which one would you choose? Perhaps a bit of each?

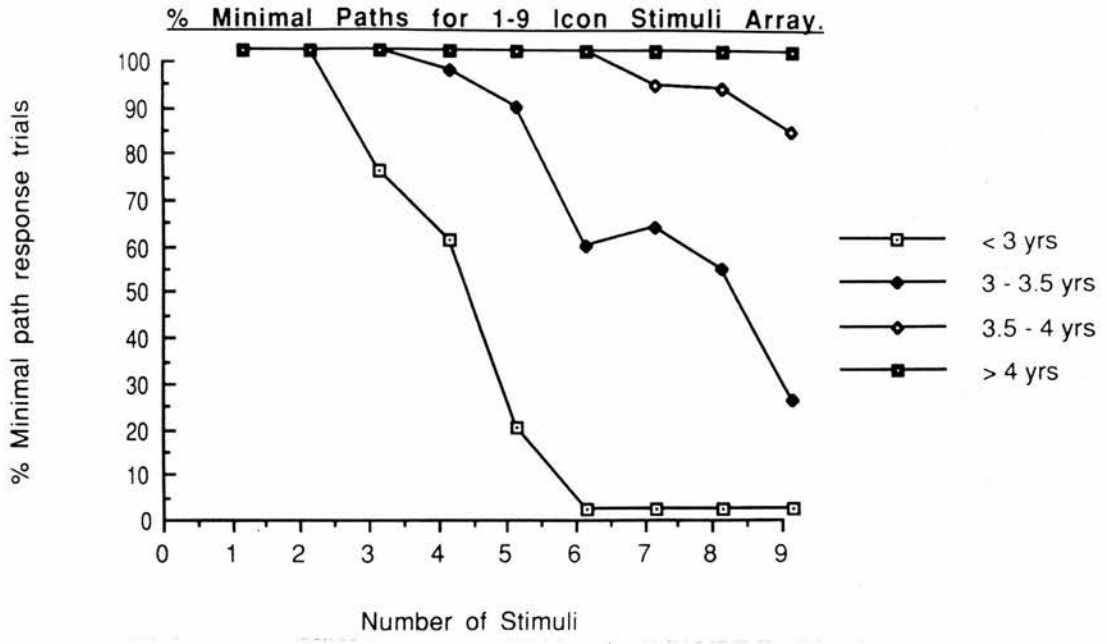


Figure 5–2: Childrens's Performance is Age Related

5.5 Characterisation of the Subjects' Performance

This description of the subjects' performance is by no means complete. It was the initial characterisation of the data for modelling purposes. The data produced from the task is extremely rich and the analysis is still ongoing. Also, the modelling presented in the next chapter was done in tandem with the empirical work, so even less was known when the modelling began. Additional analysis of the data is presented in the next chapter at the points where it informed the modelling.

5.5.1 The Childrens' Performance

Figure 5–2 shows the percentage of minimal trials on different numbers of stimuli for different ages of children. It shows clearly that the number of stimuli that a child can cope with increases as the child gets older, until the age of four, when they can cope with the maximum we can give them. Thus, we have the classic profile of a developmental task.

This raises the question: what is the difference between a child that succeeds and a child that fails. Happily, the difference is very obvious. The successful child

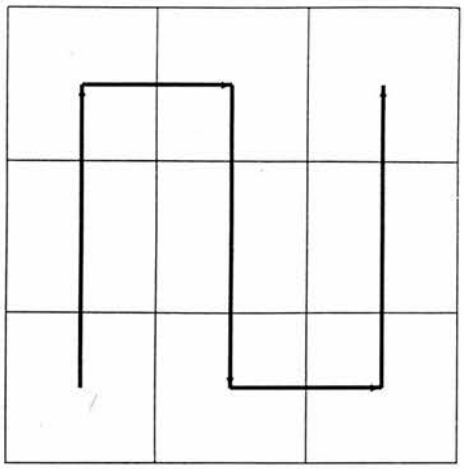


Figure 5–3: Most Common Strategy used by Successful Children

shows highly constrained trajectories through the space. Figure 5–3 shows the most common of these.

Notice that every transition is to a stimulus adjacent to the last one touched, and that there is a strong vertical vector component to the strategy. The children who failed, on the other hand, showed much less constraint in the types of transitions from one touch to the next. Jumping around in an unconstrained manner would be alright if you had a perfect memory of where you had touched. The subjects who fail, by definition, go back to touch stimuli that they have already visited before. Constraining yourself in the way that the successful children do, reduces the load on memory because it is very easy given the constraints to know exactly where you have been and therefore which stimuli remain to be touched.

Thus, the children give us a template for success, in terms of constraining possible touches in such a way as to reduce cognitive strain. It is not easy to observe the changes that lead to success in children, however, because the children who fail are fully aware of their own failure and do not enjoy it. So it is difficult to get them to practice day after day on the same task until they develop the appropriate strategies. There are no such problems with the monkeys.

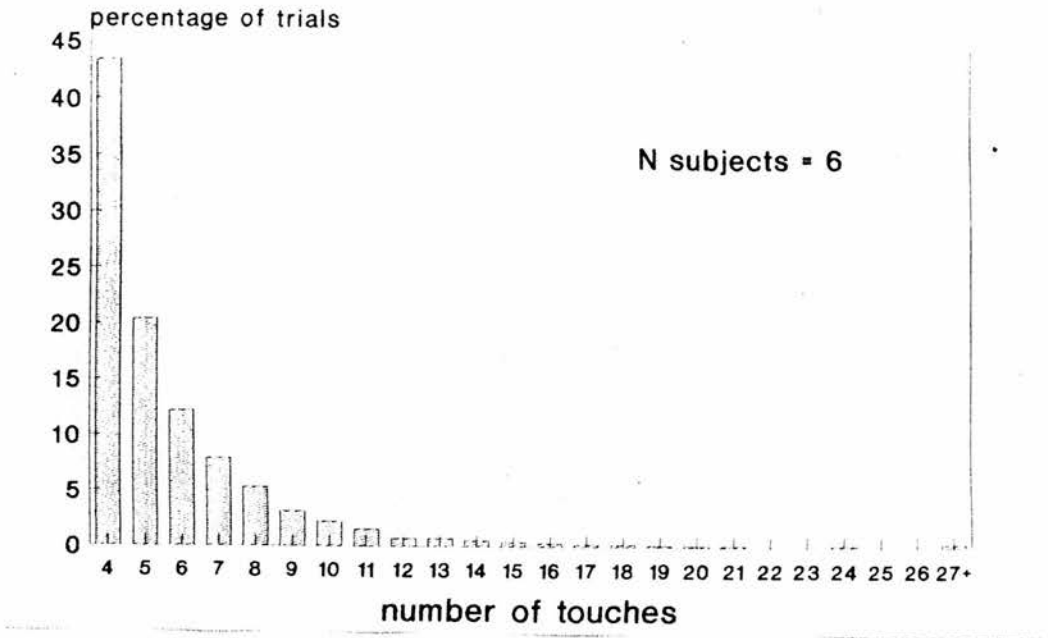


Figure 5–4: Distribution of Trials with Four Stimuli.

5.5.2 The Monkeys' Performance

Figure 5–4 shows the distribution of trials on the condition with four stimuli. The graphs for three and five stimuli, had the same characteristic shape. The modal trial length is a minimal path and there is a long tail . This tail has an improportionate effect on the average trial length making it seem as if performance is much worse than it actually is.

Figure 5–5 shows how the performance drops as the number of stimuli increases. As with the probability model, the percentage of correct touches and the percentage of minimal paths decrease at a decreasing rate as the number of stimuli increases. The performance of the monkeys, however, is significantly better than the performance of the random models. Notice that the proportion of minimal trials on four in the embedded phase is dramatically higher than the proportion in the four stimuli condition in Phase I (figure 5–4). This cannot be entirely due to task experience as the performance on four stimuli reached a plateau for each monkey before they were moved on to five. It seems therefore that exposure to higher numbers of stimuli dramatically improves performance on lower numbers of stimuli.

Figure 5–6 shows how the performance of the monkeys improves with experi-

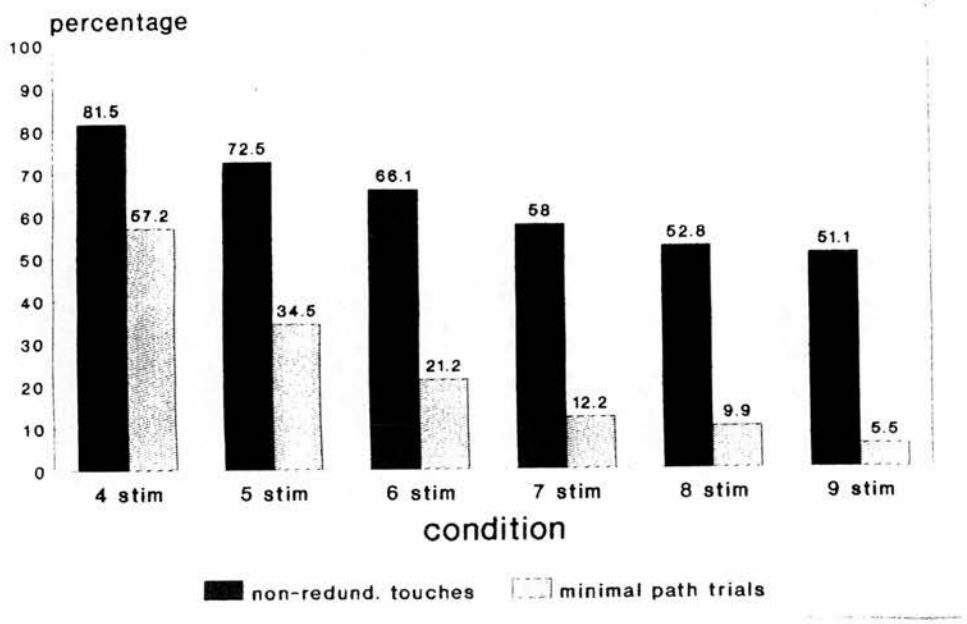


Figure 5–5: Averaged Performance on Phase II (embedded phase).

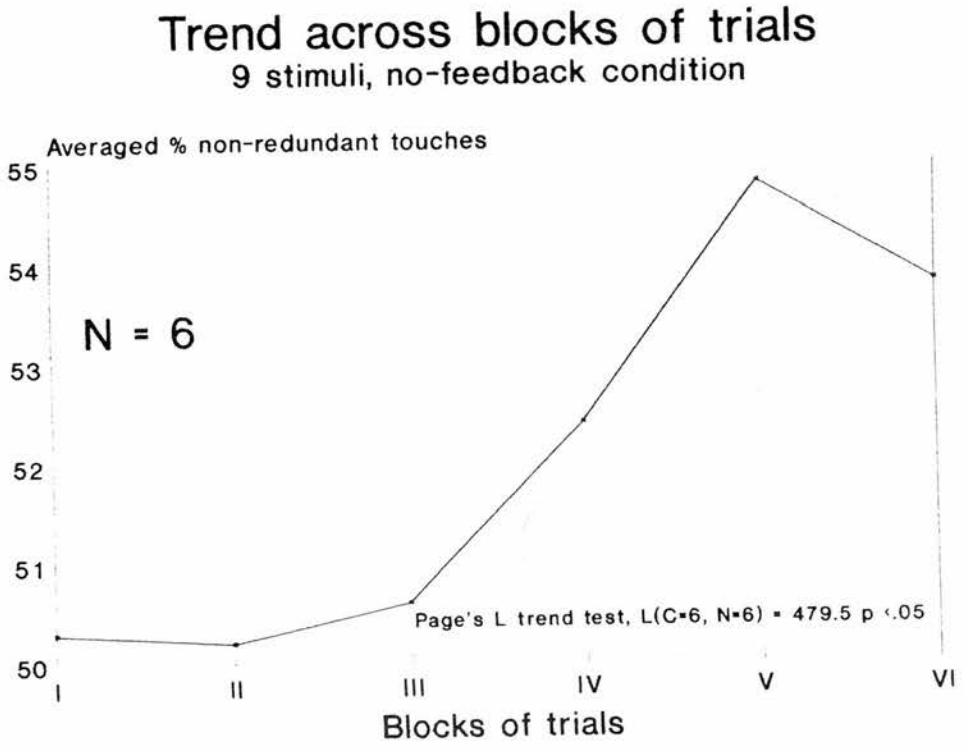


Figure 5–6: Performance in Vincent Sixths for the Monkeys on Phase III (9 Stimuli).

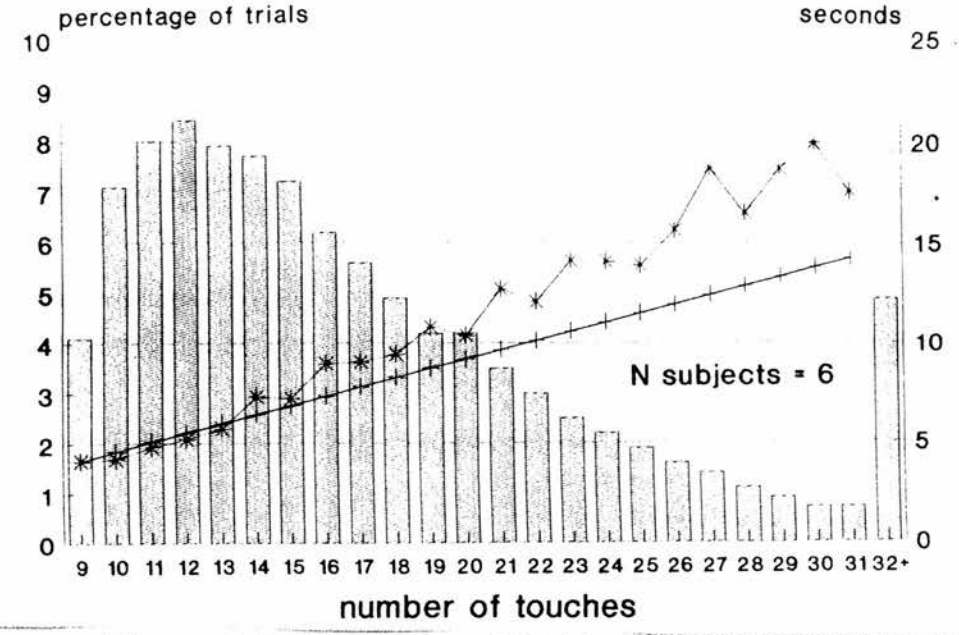
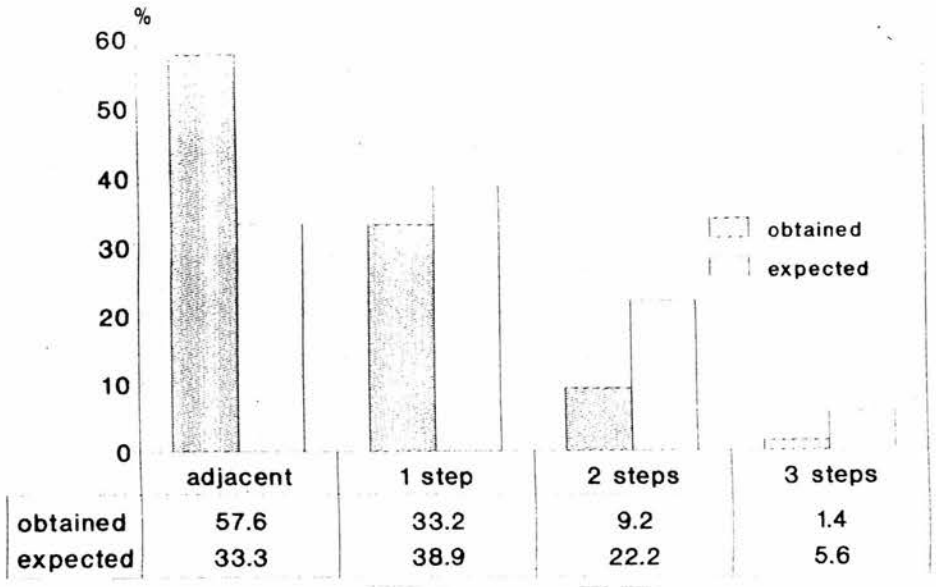


Figure 5–7: Distribution of Trial lengths on Phase III.

ence. I have shown the graph for nine stimuli as it is the clearest, but the shape is the same for three, four and five stimuli. The monkeys start at a reasonable level of performance but this performance drops slightly after a session or two. The performance then steadily improves until it reaches a plateau.

Figure 5–7 shows the distribution of trials in **Phase III** where every trial has nine stimuli to be touched. Not surprisingly the modal value is no longer the minimal path as the task is so much harder. The long tail is more prominent than ever. The two extra lines on the graph relate to the times from the first touch to the last. The straight line shows what would be predicted if each touch took the same time. The jagged line shows the averaged times that the monkeys produced. The fact that the observed response times are greater than those predicted shows that the bad trials cannot be explained by a speed-accuracy trade-off.

Figure 5–8 shows the distance of transitions in Phase III. The proportion of adjacent moves is significantly higher than chance, but still a long way off the 100% that the successful children showed.



	adjacent	1 step	2 steps	3 steps
obtained	57.6	33.2	9.2	1.4
expected	33.3	38.9	22.2	5.6

Figure 5–8: The Distance of Transitions with Nine Stimuli.

5.5.3 Summary

The performance of the children is age-related indicating that the task taps into some ability that develops. The successful children show highly constrained search trajectories with adjacency being the strongest constraint followed by a vertical vector constraint. The unsuccessful childrens' search trajectories are less constrained with a significant number of non-adjacent moves. This makes it harder for them to remember where they have been leading to them making redundant touches. The successful childrens' performance provides a template for success with which we can compare the monkeys' performance.

The monkeys perform better than chance on all numbers of stimuli upto the maximum nine. Similarly to the random models, as the number of stimuli to be searched increases, the performance on all measures drops. Exposure to higher numbers improves performance on lower numbers of stimuli. The monkeys also show a bias to touching an adjacent stimulus but it is much weaker than the childrens'. The monkeys' performance on each number of stimuli improves with experience indicating that they are regulating their behaviour.

Chapter 6

Exhaustive Search in a 3x3 Grid: the Modelling

6.1 Introduction

The modelling presented here was done simultaneously to the experimental work. This is novel. It meant that nobody knew what level of performance the monkeys would reach, nor the qualitative aspects of their behaviour, when the modelling began. This approach is not recommended if you are trying to fit the data in a *post hoc* manner with the least number of free variables! However, this approach is ideal if the purpose of the modelling is to gain insights about the cognitive workings of a task-solving agent. The main features of the model can be constantly tested against the new experimental data coming in, and because the model can be run for a huge number of trials very quickly, it is possible for the model to predict what might happen in situations as yet unencountered, and thus inform decisions about future experimental conditions. Thus the modelling and the experimental work each inform each other, and interact in a dynamic and symbiotic relationship.

This methodology does produce the problem, though, of how to describe the research. One possibility is to give a blow by blow account and another is to use all the hindsight available and so to avoid what might seem stupid now when looking back. I have opted for a compromise, where the account is basicly historical but hindsight has edited some of the details for the sake of conciseness and self-respect!

6.2 Basic Components of the Models

The models are the same sort as the transitive inference models, (Figure 3–1).

The Environment The environment includes the touchscreen and all the information on it, the peanut dispenser and all the other salient features that might impinge on the monkeys sensors. In essence, the environment embodies the task.

The Representation A large part of this modelling methodology involves experimenting with just what information the agent needs to pick up from the environment, (and its own behaviour). This is done through the representation. The representation encodes salient features from the environment that can be used to inform the decision making process.

A Set of Actions The agent impinges on the environment by performing an action. In this task the actions considered initially are touching a stimulus in a particular location in the grid. Therefore, the set of actions consists of the nine locations in the grid. For a particular trial not all locations will necessarily have a stimulus in them. For the purposes of deciding which action to perform, only the touching of locations with stimuli in them are considered possible actions and the other actions in the set are excluded from the decision making process. Actions can be elaborated in a similar fashion to the representational features. This possibility is explored later.

The Policy The agent has a representation of the state of the environment and a set of possible actions from which to choose; the policy provides the mapping from the represented state of the environment to the set of possible actions. As with the transitive inference models the policy is instantiated as a connectionist net linking the representation with the actions and there is a stochastic element to the choice process.

The Critic The agent needs to evaluate how good it's policy is, and in this task the feedback it receives from the environment is extremely poor. The critic instantiates an evaluation function for the policy, which for each state predicts the expected return from the environment which it can then compare with what is actually received and thus produce a measure of how well the policy is performing. The critic does this without knowing anything about what the policy is doing, how it does it, or what particular action was taken.

Evaluation

In this serial search task the feedback is at an absolute minimum. Most of the responses get exactly the same feedback as each other, and occasionally a response will finish the trial and a peanut will be dispensed. The agent needs to be able to evaluate responses in a much more sophisticated way than simply as to whether a particular response received a peanut. A measure that might be used to evaluate responses, and that would be more useful, would be a measure of how far (in time) the agent is from positive reward, (even if it were only a rough estimate).

In this task, let us assume that the cost of making any response is one unit and the reward of getting a peanut is also worth one unit. The unmodified value of most responses is then -1, and the value of the response that receives the peanut is 0, (-1 for the action and $+1$ for the reward). The table below shows the values of each response in a trial where the reward is received after the fourth response if the future discounted reward is used to value the responses.

Choice Number	Value
1	$-1-\gamma-\gamma^2$
2	$-1-\gamma$
3	-1
4	0

As the table shows the modified evaluation disambiguates all the values of the responses that would have been the same. The value of the first response is least (most negative), and the values increase at an increasing rate towards the goal.

6.3 Model 0: No memory

The reward for making an action that does not receive a peanut was set at -1. The reward for making an action that did receive a peanut was set at 0. The value of the discount factor, γ, was experimented with. It's exact value will be reported where it is relevant. All weights start at zero. The implicit assumption made here is that the subject starts off indifferent as to which stimuli to choose, and expects a peanut for every touch. This seems to reflect pretty well the likely state of the subject after they have been shaped to touch single stimuli presented on the screen. The models started the task at the stage where two stimuli are presented.

The representation consisted of nine input units one for each location in the grid. If a particular location had a stimulus in it then the activation of that unit was one, otherwise it was zero. There was no memory for where the last touch had been.

There is no way that this model could solve the task given it without any memory. The inputs never change during a trial and because the policy always gives the same (stochastic) response to the same input, it can only learn a superstitious ranking of preferred touches, or remain random.

The purpose of this model is to act as an ontological baseline; although it is impossible for this model to solve this task, it can solve other tasks. For example, if the experiment were a simple discrimination task where the monkey was rewarded if it touched the stimulus in a particular location, then this model could solve that task. The discrimination task could be made more complicated with the target location depending on the specific configuration. If the set of configuration-target pairs were linearly separable then this model could solve them (given the right parameter values). If the configuration-target pairs were inseparable then this model would fail.

This task being modelled is not a discrimination task. However, we are trying to get a model that does more than just succeed at exhaustive search in a 3x3 grid. To solve the exhaustive search task the model must be elaborated, but in doing so it is important not make the model incapable of solving the simple discrimination tasks that this baseline model can solve.

6.3.1 Why Space is not Initially Encoded in the Representation

The stimuli are presented in a three by three grid which highly constrains the possible search patterns. This leads to the question of whether the search patterns that emerge are a result of self-constraint or a result of the form of presentation. It is impossible to answer that question if the models are immediately given a sophisticated representation of space and its properties. The strategy here is to slowly and incrementally improve the models' representation of space such that we can then determine exactly how sophisticated the subjects representation must be and in what way.

There is also the problem of representing the spatial dimensions. It is extremely difficult, given the connectionist instantiation, to give the model an "obvious" representation of space without building in a whole load of hidden generalisations that may or may not be useful to the model. For example, suppose that we have units in the representational layer which have an activation of 1 if the stimulus they represent is in the top row, 2 if in the middle row and 3 if in the bottom row. With this representation it is not possible for the net to learn, for example, to prefer hitting stimuli in the middle row because of the monotonic nature of the activation function. This particular problem can be overcome but only by dramatically increasing the complexity of the representation. The point is that whatever representation we use, their will be hidden generalisations, (in the above example a forced generalisation from preferring the middle row over the bottom row to preferring the top row to the middle row). By starting with no spacial representation whatsoever, and then building it up slowly in a way that is most

constrained by what the subjects actually do, we not only keep the models as simple as possible, mitigate this inherent problem to a maximum, but also learn more about what is the minimum spatial representational resource required.

6.4 Model 1: Remembering the Last Location Touched

If the agent knows which position it touched last, it has the potential at least to solve the problem. To the nine input units of **Model 0** denoting the configuration, were added nine more units, (one for each location). The unit corresponding to the location last touched has an activation of one; all the others have an activation of zero.

The model learns to avoid hitting the last location touched. This manifests itself in strong negative weights between the memory units and their corresponding action units. This is learned during the phase with two stimuli. The model fails on three stimuli. It gets stuck in loops going round and round. It is interesting to note that both **Model 0** and **Model 1**, start by choosing locations to touch at random, yet through trying to learn to perform better, they actually perform worse than the random models.

Table 5–2 in the previous chapter constitutes an optimal policy for making touches. The model can represent this table so its failure to move beyond trials with 3 items is not because it was impossible to represent the solution. It must therefore be because it could not learn the solution. To determine whether this is the case we can examine the inductive space and specifically the error space.

Learning a set of transition preference rankings

The real subjects are first given one stimuli then two and so on up to nine. With one stimuli, every possible set of rankings is just as effective as any other, so the model can learn nothing at all from the one stimulus condition. With two stimuli,

it is only necessary to learn not to repeat. The model learns a "lose→shift" policy and because there is only one other place to shift to this works optimally. When the number of stimuli reaches three, however, the model crashes. It starts off reasonably well; it hits one of the stimuli and then because it has learned not to repeat it hits another. It will not repeat the last choice so it either hits the third stimulus and does a minimal trial or reiterates back to the first stimulus it touched. So it begins doing fifty percent minimum trials on three, but this performance rapidly decays. The reason for this is obvious when you realise that to perform perfectly it has to have the first seven rows of table 5–2, or one of the alternative correct tables, fully instantiated.

The table assumes an absolute ranking when the ranking might be stochastic. However, there is every reason to believe that a stochastic ranking would be worse except of course when comparing a totally random model to one stuck in a loop.

The problem of looping is catastrophic. It arises because the values given to actions are based totally on relative values, and because the net can get into a depressed state. What happens is that when an action receives a reward which is less than expected, the agent is less likely to make that action in that situation again, but it also reduces its expectation of reward for that situation. If this happens enough its expectation of reward becomes so low that it starts to accept its own low level of performance. Therefore, the net gets into a situation where it keeps reiterating forever, and expects never to receive a peanut. It is impossible to imagine a monkey getting into this state. For a start if it stopped expecting peanuts it would stop responding at all, whereas the model has no choice in the matter. There is a sense in which receiving no reward ever is an absolute minimum. With the rewards set at -1 and 0 for no peanut and peanut respectively, the expected discounted reward tends to

$$\frac{-1}{1-\gamma}$$

and this value could be used as a trigger to do something drastic to the policy; afterall nothing could be done that could make performance worse. Well, that is not necessarily the case. The agent does not necessarily know that changing its

policy might not lead to huge electric shocks, but considering that in it's experience of the task to date, it has only received no reward or a peanut, it seems impractical and negative to get stuck in such a rut.

The model starts the three stimuli condition knowing only that it should not repeat and otherwise moving pretty randomly. This works pretty well, but the model (already having to lower its expectancies because now at best it is getting a reward after every three choices rather then every two) wants to do better. It tries to learn a positive strategy which involves fixing some of the transitions. As the vast majority of the fixing produces a loop the performance drops rapidly.

6.4.1 How the Agent Evaluates its Performance

To understand how a series might be learned by the system it is necessary to understand how the critic evaluates each step in that series. The crucial thing to remember is that the critic does not evaluate series but specific actions.

At first, the series produced are more or less random and the reward received for each action is usually the same. The critic learns what this value is and thus starts predicting that future rewards will be similar. Occasionally, an action will receive a reward greater than expected, because it will be at the end of a specific sequence that has become an exhaustive search. The initial randomness of choice coupled with the small number of stimuli (the maximum number of stimuli being nine) ensure that eventually there will be some differential feedback. When this happens, the critic adjusts its evaluation for *this action taken in this state* and reinforces the connections between the state and the action thus making it more likely in the future for that action to occur when the system is in that state (or to a lesser extent a similar state).

The extent to which these changes improve the overall performance will depend critically on the representation of this state and the extent to which it motivated the action taken. Given that the state is represented only by the set of stimulus positions in the current configuration and the last stimulus touched, the improvement may be non-existent or even detrimental. The configuration is present for

the whole trial and so cannot usefully inform the policy as to which action is best. The agent soon realises that repeating is bad, but it cannot learn more as the number of stimuli rises past three.

6.5 Model 2: Adding More Memory

By increasing the amount of working memory it should be possible to improve the models performance as the negative strategy is the easiest to learn. The hope is that once the negative strategy becomes sufficiently strong a positive strategy can also emerge. The obvious way of doing this would be to have another set of nine units that would encode the position touched two time steps ago. This would require nine extra units for not a great deal of extra memory. Furthermore, it seems a rather simplistic notion of working memory.

Another possibility is to keep the same number of units for working memory but have the activation of these units decay over time rather than just vanish immediately. In other words the traces in memory for the positions that have been touched fade slowly away with time.

Model 2 has nine units that record the last response made; the same as **Model 1**. The activation of these units, however, is reduced by multiplying the activation value by the value of a decay parameter, δ, at the end of each time step. The activation of the unit corresponding to the last response is incremented by one at each time step.

By having the trace of the response decay without having any distinguishing features means that there is a forced generalisation from learning not to repeat immediately to learning not to reiterate. The extent of the generalisation depends on the value of the decay variable. There is also generalisation from the positive transitions that interfere with the non-reiterating part of the policy. Suppose for example, that having hit the stimulus in location 3, the policy states that it should not hit 3 again and (all other things being equal) should hit the stimulus in position 4. At the next time step, there will be a strong bias to not hitting 4,

a weaker bias to not hitting 3, but there will also be a weak bias towards hitting 4. One result of this is that if a stimulus is touched for no reason, i.e. it was an exploratory touch, then it is likely to be avoided more strongly than if there had been a strong weight producing that touch.

6.5.1 Performance of Model 2

As the discount factor, γ and the decay parameter, δ, increase the model can cope with increasingly higher numbers of stimuli. However, as these parameters get higher, the ability to cope with a particular number of stimuli becomes more fragile; i.e. the higher the values of the two parameters the greater the number of stimuli that it can possibly cope with, but the lower the probability of reaching those numbers without getting stuck at a lower level. With γ and δ both equal to 0.75 the model is near to its optimal performance as measured by the number of stimuli it can cope with and its robustness against depression. With these values the model shows performance levels similar to the monkeys upto six stimuli. Once the number of stimuli increases above six the model tends to crash into looping and depressed behaviour.

It is important and interesting to realise that adding memory *ad infinitum* in this way does not help. Omniscience without omnipotence is of little use in this task.

The model showed a number of characteristic features in common with the monkeys performance. It succeeded in reaching the criterion for two, three, four and five stimuli. The number of sessions required to reach criterion tended to increase as the number of stimuli increased. When it was advanced onto a new number of stimuli, the model performed well at first, then dropped sharply before recovering up to a stable plateau. Finally, when the net is trained on five stimuli for a couple of sessions and then moves back to four stimuli, there was a dramatic improvement in performance on four stimuli. Of course, the model fails to cope with numbers of stimuli higher than six and neither does it show any adjacency bias.

Analysis of the policy

Examining the weights in the policy it is possible to distinguish three separate components within it.

1. Aversion: The model learns to avoid touching a stimulus it has already touched. This can only work to the extent that the model can remember where it has touched. So increasing the decay parameter(which reduces the decay over time) increases the number of stimuli that the model can cope with, but as mentioned before this is limited.

2. Starting Points: At the beginning of a new trial there is no previous response to influence which location should be hit. Therefore, only the inputs denoting the configuration will affect the policy. The overall effect is that each configuration will produce a ranking of the locations present in terms of how preferable they are to touch. As the first touch depends entirely on these preferences they can be considered as starting point preferences.

3. Transitions: Preferred transitions appear whereby having touched a stimulus in one position, there is a rank of preferences for touching a stimulus in another position.

The transitions and starting point preferences interact in a positive way with each other and the aversion part of the policy. For example, in one case the aversion for retouching the stimulus at position 4 in the grid was considerably weaker than for retouching any of the other positions. The aversion was still strong enough to avoid an immediate repetition or a reiteration after just one other response but no more. This weakness did not impair performance, however, because the stimulus at position 4 was never touched too soon in the sequence. Thus we find that touching position 4 was never a preferred starting point or even near the top of the rank of preferred starting points. Furthermore, the transition preferences support this, by ensuring that position 4 will get touched but not immediately. By starting in particular positions the agent can also avoid an inconsistency in

the transitions. If there are two inconsistencies in the transition preferences then starting points are of little use.

It is difficult to disentangle cause and effect when analysing the interactions of different parts of the policy, but the example above does illustrate how the self-regulation inherent in the model can cause the policy to organize itself in a way to produce useful interactions between its different component strategies.

Performance profile within a condition

When the model starts on a new number of stimuli it performs quite well to begin with. This is because is has largely consistent transitions, but rates its own performance badly because it still expects to be rewarded after $n - 1$ touches even though there are now n stimuli. Thus every time it makes a touch the reinforcement tends to be negative, making it less likely to go back to that position. Unfortunately this honeymoon period cannot last - the good set of transitions it previously had are slowly unlearned (though not completely) and the critic soon learns to expect less overall reward. The agent then goes through a very bad period when the trial lengths increase dramatically before improving again upto a plateau.

Retraining on 4 stimuli after exposure to 5

A very obvious effect that arises from both the model and the performance of the monkeys is that exposure to a higher number of stimuli improves performance at a lower number. Analysis of the policies before and after the exposure reveal the following differences:

1. The aversive part of the policy has become even more aversive. This would lead to an improvement because although the memory of previous responses remains at the same level throughout, it allows the use of fainter traces to become significant in avoiding reiterations.

2. The starting points become stronger and many more configurations will share the same starting point. In other words, before exposure to greater numbers of stimuli, the starting point preferences were very much bound to the specific configuration that was being presented at the time, and after exposure the starting point preferences were for large groups of configurations.

3. The transition preferences become more internally consistent, i.e. give rise to larger rather than smaller loops.

6.5.2 Similarities and differences with search in Artificial Intelligence(AI)

There are striking similarities between the decomposition of searches presented above, and those found described in any undergraduate AI textbook. This is no coincidence and it is worthwhile to consider the reasons behind these similarities and to note some important differences.

Any problem which can be defined in terms of an initial state, a goal state, and operators or transitions which change the state, is basically a search problem. A huge number of problems, including planning and scheduling problems, can be defined in this way which is why search plays such a central role in AI. Search problems in general are computationally intractable – as the number of states increases linearly, the number of possible searches increases exponentially. Furthermore, it is often found that the expected case complexity of a search problem is very close to the worst case complexity. This creates the need for a large amount of guesswork. Guesswork – the use of heuristics – attempts to make the search problem tractable by reducing the search space as much as possible. It is this characteristic of search problems that motivated the psychological investigation presented here – in order for a developing system to adapt yet remain adaptive it must constrain its behaviour appropriately in the face of a huge number of possibilities.

So the motivations for looking at search are different for the two disciplines. AI is interested in search because a large number of tasks can be neatly defined as

search tasks and therefore if you can solve the problems inherent to search tasks you have solved a huge number of tasks. We psychologists, on the other hand, came to look at search tasks because the problems that are inherent to search tasks require subjects to self-constrain their behaviour in a way that will shed light on the major problem of cognitive development itself.

Given that the key to efficient search is to reduce the space of possible searches as much as possible, then starting points, preferred transitions, and aversion to returning to states that have already been visited, fall out naturally as easy and efficient constraints. What the experimental programme has added, however, is empirical evidence that real agents can apply these different types of constraints. Of course, we are also suggesting that subjects apply these constraints precisely because they are efficient and easy – easy on resources and easy to implement.

Fixed starting points reduce the number of possible searches by a factor of the number of items to be searched. They also give the advantage of firmly anchoring a particular search sequence. The psychological significance of this is huge as has been shown in transitive inference tasks for example, (McGonigle and Chalmers 1986). The importance of this end anchor is that it provides the subject with a fixed point both with which to attach other items and to return to should the sequence run into problems. It is not easy to show statistically that monkeys used starting points in this way, but certainly observation gave us the impression that often a subject would start a sequence at a particular location, make a number of non-reiterating touches fairly rapidly, and then "get lost", at which time reiterations would be made. At this point the subject would go back to the starting point and quickly produce an optimal search. The total trial length of these sorts of trials is unimpressive although the second half of the trial is optimal in itself. Thus, a fixed starting point is very useful not only for dramatically reducing the dimensionality of the search space, but also as a key part of a recovery procedure.

This aspect of starting points corresponds loosely to backtracking in AI search algorithms. Consider a depth-first search of some tree structure. When the search reaches a terminal point which is not the goal, a classic AI search procedure will backtrack to the last choice point and take the next possible route. Of course,

this exhaustive search task does not fit nicely into a well ordered tree structure, because with looping and reiterations possible there are no terminal points. It is not surprising therefore that the recovery procedure should go right back to the beginning and start again.

One must remember that subjects are not told that the task is an exhaustive search task and that they therefore have to work out for themselves that reiterations and repeats are bad. This is the reason that the search space cannot be neatly structured into a nice tree. This highlights perhaps the biggest difference between the psychological and AI approaches. AI has long realised that returning to a repeated state is a major problem in search tasks. State duplication leads to an exponential degree of redundancy, and unless there is some means to deal with this redundancy we can get into catastrophic looping. This is where the aversive part of the policy plays a crucial role by discouraging returning to a repeated position.

In the models presented here working memory is a limited resource and aversion to repeated states must be learned. This is the only way to have a psychologically plausible model that does not assume that the agent already knows the nature of the task. In AI, on the other hand, psychological plausibility is not important, and only rarely are AI practitioners wary of building in task knowledge. Thus the classical AI strategy to avoid the repeated states problem is to store every single state that has ever been visited. Research then has concentrated on how best to represent this potentially huge amount of information to make storage and retrieval as efficient as possible. The most famous example of this is the "discrimination tree or net" pioneered by Newell, Shaw and Simon(1958) and used in GPS.

6.5.3 How Constraints Lead to the Unrepresentability of the Correct Evaluations

This subsection presents a *reductio ad absurdum* proof that the model as it is could not possibly represent the correct evaluations for an optimal policy.

Consider the task when there are four stimuli, and a policy that produces minimal paths every trial for every configuration. The value of a particular state with respect to the optimal policy depends totally on the current sequence position. If the reward for making a touch and receiving no peanut is -1 and the reward for making a touch and receiving a peanut is 0, then the values for the different sequence positions are as follows:

Choice Number	Value
1	$-1-\gamma-\gamma^2$
2	$-1-\gamma$
3	-1
4	0

For the first choice only the input units denoting the configuration have any activation, therefore only they can contribute to the evaluation prediction of the first choice. Any four of the nine locations can have a stimulus in it for a specific trial. As the evaluation of each configuration must be the same for the optimal policy being considered, the weights between every input denoting part of the configuration, and the critic must be equal to each other.

Furthermore, no more than two locations can have a single sequence position for every configuration; the policy may aways hit the stimulus in location 1 first, when it is present and always hit the stimulus in location 9 last, when it is present, but if it does this then all other locations will have at least 2 sequence positions.

Remember that the value given to an action is given by:

$$V_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots + \gamma^{t-1} r_{t+n} + \cdots$$

where the reward r is -1 if no peanut is received, else it is 0. Considering the

evaluation function of an optimal policy for n stimuli we get:

$$\begin{aligned}
V_1 &= -1 - \gamma - \gamma^2 - \gamma^3 - \cdots - \gamma^{n-3} - \gamma^{n-2} \\
V_2 &= -1 - \gamma - \gamma^2 - \gamma^3 - \cdots - \gamma^{n-3} \\
&\vdots \\
V_{n-3} &= -1 - \gamma - \gamma^2 \\
V_{n-2} &= -1 - \gamma \\
V_{n-1} &= -1 \\
V_n &= 0
\end{aligned}$$

The value, V_1, must be totally determined by the configuration units which do not change during a trial. The difference between V_1 and V_t, for $n \leq t > 1$, must therefore be given by the other units. Considering **Model 2**, the difference between V_1 and V_t, for $n \leq t > 1$, must be given by the memory units. Specific grid locations cannot be associated with specific sequence positions, therefore, all the locations should contribute equally to the evaluation; i.e. the weights between the memory units and the critic should all be equal to each other. Let the value of this weight be denoted by w. Let M_t = the total activation from the memory units at time, t. We find then that

$$\begin{aligned}
M_2.w &= V_1 - V_2 &&= -\gamma^{n-2} \\
M_3.w &= V_1 - V_3 &&= -\gamma^{n-3} - \gamma^{n-2} \\
&\vdots \\
M_t.w &= V_1 - V_t &&= -\gamma^{n-t} - \gamma^{n-t+1} - \cdots - \gamma^{n-3} - \gamma^{n-2} \\
&\vdots \\
M_n.w &= V_1 - V_n &&= -\gamma^0 - \gamma^1 - \cdots - \gamma^{n-3} - \gamma^{n-2} = V_1
\end{aligned}$$

Model 2 had a decay function on the activation of the memory units and then incremented activation of the unit corresponding the the location last touched by one. The crucial question is whether by altering the decay parameter it is possible to get the critic to produce the correct evaluations.

So $M_2.w = -\gamma^{n-2}$ but M_2 must equal 1 because the decay function has not had time to work yet. Therefore $w = \gamma^{n-2}$. Let the decay function be denoted by

$f(x)$ such that if $M_t = x$ then before incrementing $M_{t+1} = f(x)$. Let $f^y(x)$ denote applying the function f to the value x, y times. Then

$$\begin{aligned}
M_2 &= f^0(1) && = 1 \\
M_3 &= f^0(1) + f^1(1) && = 1 + 1/\gamma \\
M_4 &= f^0(1) + f^1(1) + f^2(1) && = 1 + 1/\gamma + 1/\gamma^2 \\
&\vdots \\
M_t &= f^0(1) + f^1(1) + \cdots + f^{t-2}(1) && = 1 + 1/\gamma + 1/\gamma^2 + \cdots + 1/\gamma^{t-2}
\end{aligned}$$

It is quite obvious from this that the decay factor, δ, must equal $1/\gamma$, for the correct evaluations to be representable. But this makes no sense: if they are not both equal to 1 then either γ or δ must be greater than one, which means that either the agent should value rewards in the future more than rewards in the present, or it should remember the distant past better than the recent past!

With γ between 0.0 and 1.0 and decay in the same range, the critic cannot represent the optimal evaluation function for the optimal policy for stimuli numbers greater than two. What happens is that the critic underestimates the value of the first and last choice and overestimates the value of the middle choices; the errors tend to cancel themselves out this way. If units have an effect over all choice positions then the effects will cancel out. Thus, in **Model 2**, all the transitions tend to be reduced but because they work in a relative manner their effect still operates. It is a credit to **Model 2** that despite the fact that it cannot represent the correct evaluations for the optimal policy it still manages to do so well.

To solve this problem it is necessary either to change the way that the evaluation works or to elaborate the representation still further. The former might be done by introducing hidden units between the representation and the critic and altering the activation function and learning rule to a sigmoidal function and back-propagation respectively. There are a couple of problems associated with this approach though.

The transitive inference modelling showed that it is important for the critic to lead the policy. The introduction of hidden units into the critic will necessarily slow down the learning of the evaluations, which would make it virtually impossible

for the critic to lead. Whether it worked or not, it would make analysis of the critic substantially harder, and this goes against the philosophy of the elaborative modelling strategy.

There may be other ways to change the critic but I have not found them, which leaves us with the possibility of elaborating the representation such that the evaluations can be generated easily. There are many ways to elaborate the representation but the simplest and surest is to change the instantiation of working memory.

6.6 Model 3: Changing Working Memory

Rather than having one set of units to cover the whole of working memory, as above the model was changed so that the location touched last was remembered on one set of units, the location touched second on another and so on. Obviously, this produces a potentially infinite number of units, but this can be avoided by maintaining a decay function and so in effect, only five or six sets of units actively contribute to performance, depending on the value of the decay parameter. This model is obviously less parsimonious than the previous models but at least now the model has the potential to represent the correct evaluations on optimal performance to a certain degree of accuracy.

The performance of **Model 3** is worse than the previous ones. With the same parameter values as **Model 2** it usually fails when the number of stimuli is one less. One reason is that there is no enforced generalisation from learning not to repeat to learning not to reiterate. The model does learn that reiteration is bad but it takes much longer and the effect is not so strong. The evaluations are very similar to those of **Model 2**: i.e. the first and last touches tend to be undervalued and middle ones overvalued, so the problem with the critic is not merely to do with computability but also to do with learnability.

6.7 Three Ways of Encoding Adjacency

The basic search model with working memory units, whose activations decay with each time step, can learn to do the task with a similar economy to the monkeys upto five items (more or less, depending on the specific values of the discount factor, γ, and the decay parameter, δ). The pattern of responses is not similar however, because once the number of stimuli reaches four the monkeys show a strong adjacency principle; the model cannot possibly show such an adjacency principle except by pure chance, because it has no notion of how far apart particular stimuli are from each other. The **Monte Carlo Random Walk Model** shows that adjacency on its own is no good, but when it is combined with the aversive part of the policy and the positive transitions this may change. What might be hoped is that the aversive part of the policy gets rid of the most detrimental aspect of adjacency, namely that when two stimuli are adjacent to each other but a jump away from any others, then adjacency can lead to reiteration over the two adjacent stimuli and neglect of the others. Thus the hypothesis is that adjacency will lead to the exhaustion of sub-areas of the grid, the negative strategy will ensure that the agent jumps out of that area after exhaustion, and the positive transitions will only have to deal with moving from one sub-area to another, which is much easier because there are far fewer sub-areas than there are stimuli. These sub-areas are not built into the representation; the hope is that they become an emergent property from the interaction of the different components of the policy.

6.7.1 Adding Adjacency Units to the Representation

In this model units were added to the representation. The new adjacency units encoded which locations with stimuli in them, were adjacent to the stimulus last touched. There were nine extra input units for doing this, one for each location. For the first choice of each trial, all of the units had an activation of zero. For subsequent choices the activation of an adjacency unit was one if the corresponding

location contained a stimuli, and the location was adjacent to the stimulus last touched. Otherwise it was zero.

The performance of this model was always slightly worse than the performance of the base model. The agent learnt to use adjacency on trials with two stimuli. Obviously, when there are only two stimuli, moving to an adjacent stimulus means not repeating, so it will be reinforced. Once the number of stimuli increases above two, the initial bias is overturned such that there is a *negative* bias to moving to an adjacent stimulus.

Adjacency and 3 stimuli

With 3 stimuli and no constraints other than adjacency the model learns not to move to an adjacent item. The reason for this is that a strong adjacency bias does lead to the neglect of one of the stimuli when the other two are adjacent to each other. However, if there is a positive bias *not* to move adjacently, performance will not be too bad on any of the possible configurations. With greater numbers of stimuli, the jumping policy has a negative effect on performance.

6.7.2 Adding an Adjacent Unit to the Actions

The output units can be elaborated with a further unit whose activation is automatically added to outputs of adjacent units. As with the other units the adjacency unit would learn only when an adjacent move was made.

The model still learns (with three stimuli) that adjacency is bad, and it can never unlearn this with higher numbers because adjacent moves are already discouraged and so do not take place. Therefore there is no learning.

6.7.3 Changing the Rewards

The only way to make sure that there is an adjacency bias regardless of what the agent has learned, is to incorporate it into the rewards. At the moment the reward

	Starting Positions in Grid									
Subject	1	2	3	4	5	6	7	8	9	χ^2
Charlie	15	37	59	74	86	84	165	90	31	219.5
Alfie	101	35	30	178	79	35	223	92	27	442.0
Ollie	84	72	81	87	106	115	42	75	78	42.8
Kissy	106	44	16	95	18	18	39	8	12	271.7
Mimi	6	59	429	4	35	140		2	14	2035.8
Luba	62	73	29	97	102	42	82	50	23	104.5

Table 6–1: Starting Point Frequencies on Phase III

for most actions is -1; in subsequent models this value is modified according to the distance of transition by subtracting the distance measured by adjacent moves multiplied by a constant, α. When $\alpha = 0.1$ then the reward for a repeat is -1, the reward for an adjacent move is -1.1, the reward for a move of distance 2 is -1.2 and so on.

It is not possible to determine empirically whether the monkeys are using the number of touches or the time till the next reward as the basis for their evaluation, because there is no independent third measure of performance to compare each against, and anyway the two measures are very highly correlated so it would be difficult even if there was. If it is accepted that the subjects may be using time instead of the number of moves, then this way of implementing an adjacency bias is not totally unprincipled, as the subjects do make adjacent moves faster.

When the rewards are modified in this way an adjacency bias comparable to the monkeys does arise, but it does not improve the performance of the model in any way.

6.8 Starting Points

Table 6–1 shows the monkeys starting point preferences on **Phase III**. These are significantly difference from chance, (χ^2 test, eight degrees of freedom, $p < 0.01$), even Ollie's. The starting point preferences in this phase are stronger than on previous phases with fewer stimuli, and this would be expected because when there are fewer stimuli, a preferred starting point in the grid may not contain a stimulus. Despite the significance of the χ^2, the preferences do not seem particularly strong except for Mimi perhaps. But having a fixed starting point would seem to be an obviously useful constraint. If you always start in the same position there is one less transition to be worried about.

The models show stronger starting point preferences early on but these tend to diminish by the time the number of stimuli reaches five. The starting point preferences do not merely disappear so that the starting positions are random. There is large positive reinforcement for the last touch whatever that location of the stimulus touched. This makes the last touch more likely to be touched earlier in the sequence and this produces a cycling of preferred starting points.

6.8.1 Specialised Starting Point Units

It is possible to add specialised units with their own evaluator to choose a starting point. Nine units were added, one for each grid location. The activations of these units are zero unless it is the first touch of a trial, otherwise their activations are given by a single weight for each unit. These units have a special evaluator that predicts the length of the trial. At the end of the trial, the unit corresponding to the starting point actually chosen is updated as is the starting point evaluation. The error for the evaluator is just the difference between the predicted trial length and the actual trial length, and the sign is reversed to get the error for the starting point units. This means that if a starting point leads to a trial that is shorter than predicted, the agent will be more likely to start in that position again.

This mechanism does produce strong starting points. However, there are two problems with it. Firstly, as there is still some stochastic element to the first choice, occasionally the model will start at the non-preferred position. As the agent has learned by this time never to go back to the preferred starting point, when this happens the preferred starting point is always neglected. Secondly, if the previous problem is overcome, by, for example, putting a limit on the stochastic element of the first choice, the model has starting points which are totally fixed whereas the monkeys show much weaker preferences.

The strong starting point preferences do slightly help the performance of the model but still the monkeys perform much better on the higher numbers. If the agent has a fixed starting point, then the number of possible sequences, although still huge, is nine times less huge. This reducing of the number of degrees of freedom in performance helps the agent but they need to be reduced much further.

6.9 Other Constraints

There are many possible ways of reducing the degrees of freedom of the agent. However, the aim is that the agent should be able to limit its *own* degrees of freedom. There may well be some complex resource management module within real agents that does something along these lines, but that is beyond the scope of this thesis. A cheaper version can be bought in terms of elaborating the output units. It seems paradoxical to add extra possible actions in order to reduce the degrees of freedom, however when the actions that are added are merely principled categorisations of the existing actions, no 'new' actions are really being added. The agent now has the opportunity of using fewer of the actions available to it because some of those actions do different things in different situations. For example, the adjacency action, mentioned earlier, does not cause the agent to touch the same stimulus every time it chooses this action. Therefore the agent can use this action appropriately more than once during a trial. Adjacency on its own does not improve performance but if an adjacency bias were working alongside other

Location Just Hit	1	2	3	4	5	6	7	8	9
Location Next Hit: 1	7	509	38	373	32	5	97	2	
2	86	23	584	119	91	54	44	13	5
3	171	47	18	45	214	245	11	45	9
4	92	31	4	35	304	9	555	74	2
5	56	62	14	37	10	319	92	271	66
6	104	51	54	77	32	36	26	158	315
7	98	40	5	213	74	7	19	407	6
8	126	75	31	62	102	43	44	13	494
9	248	139	47	194	30	143	86	32	11
column total	988	977	795	1155	889	861	974	1015	908

Table 6–2: Charlie's Transition Frequencies on Phase III

constraints such as a vector principle, this might change. So what other constraints might we consider?

6.9.1 Learning from Charlie

Charlie was the best monkey at the end of **Phase III**. Analysing more closely what he was doing suggests further constraints.

Table 6–2 shows Charlie's transition frequencies for all sequence positions on **Phase III**. Only one possible transition, (from position 9 to position 1), was never taken, but there is a large degree of differentiation between the different transitions.

To present the information in table 6–2 in a more easily interpretable form, it is necessary to abstract the key elements of the table. All the transitions which account for more than one eighth of the transitions *from* a particular position are represented by arrows in Figure 6–1. The size of the arrows increases with the frequency of the transition and adjacent and non-adjacent transitions are shown separately to keep the diagrams as clear as possible.

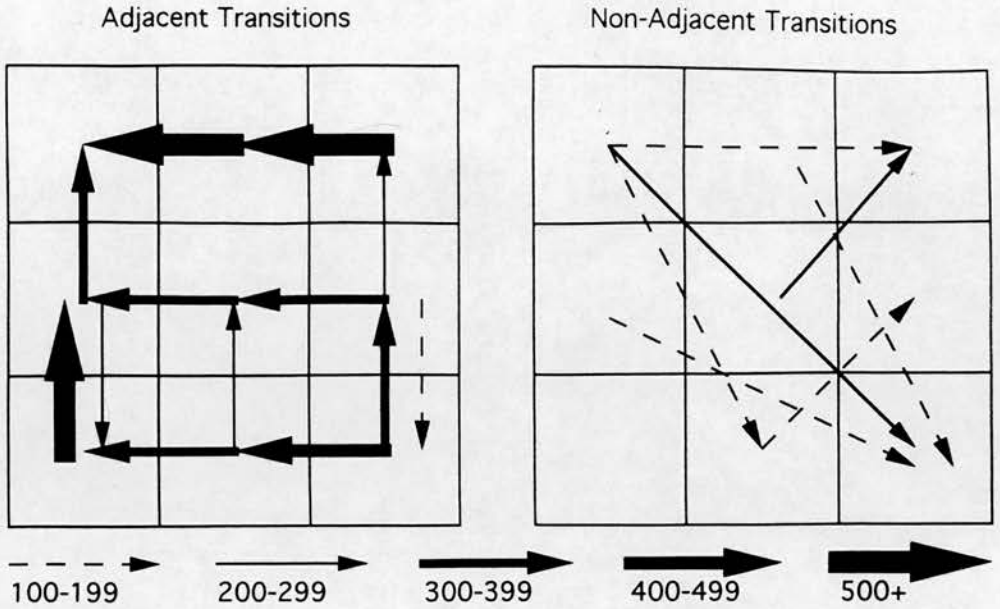


Figure 6–1: Diagramatic Representation of Charlie's Transitions in Phase III

Firstly notice from Figure 6–1 that the adjacent transitions are much stronger than the non-adjacent ones. Secondly, there are no adjacent moves to the right, but every non-adjacent move is to the right. There is also a strong tendency in the adjacent moves to move up, whereas most of the jumps are down as well as right. The only two non-adjacent moves that go up are the diagonal moves from positions 8 to 6, and 5 to 3. These reflect one of Charlie's most common minimal paths: $7 \rightarrow 4 \rightarrow 1 \rightarrow 9 \rightarrow 8 \rightarrow 6 \rightarrow 5 \rightarrow 3 \rightarrow 2$. This strategy exhausts the left hand column jumps to the bottom right hand corner and then zigzags up the central and right hand columns. Charlie produced this sequence on many times which raises the questions: why does Charlie use this strategy rather than, say, the one most children used, and why did it not become more solid such that he produced it every trial? Figure 6–1 provides the clue. It seems clear that Charlie is trying to use the same sort of transition for each move: "Move adjacently up or left". This strategy naturally moves you to position 1 where it is no longer applicable, so Charlie jumps down and right. So maybe the reason Charlie does not come up with the childrens strategy is because it involves four adjacent up moves with two adjacent down moves in the middle. The generalisation from what Charlie does in one position to what Charlie does in others is further shown by the jumps from positions 2 and 4 to 9. The two diagonal moves are essential for

this sort of strategy to work but they do not fit into the same category of moves; they are specific transitions from specific squares. Therefore we might consider the reason that Charlie was unable to maintain and solidify this strategy was because usually he could not resist moving from position 8 to 7 or 5 to 4.

Thus it seems as if Charlie was following a very strong data-reducing strategy but unfortunately the strategy does not work particularly well. Four of the other five monkeys had very similar strategies to Charlie although they were not so strong. The odd monkey out was Alfie who had a rotational strategy. He went up the left hand side, right along the top, down the righthand side, and left along the bottom. His problem was that he couldn't work out when to go to the middle. If you go slightly too late then you still produce an efficient if not optimal search, but if you go too early, as Alfie tended to, then you are in trouble. Alfie did not go back to where he had entered the centre and continue round but picked a new point on the perimeter to continue his rotation. This coupled with the fact that he was in so much of a hurry that he often jumped one of the centre side positions and cut corners, led to his performance being one of the worst. Eventually, even Alfie's strategy changed to one similar to Charlie's. In Alfie's defense it must be stated that he received peanuts quicker than any other subject because he could rattle out a trial of twenty touches faster that the others could produce a minimal path!

This rotational strategy might be considered even more data-reducing than the alternatives, and was observed in a couple of the successful children. If you start in the bottom lefthand corner pointing up, then the strategy can be represented as "straight straight right, straight right, straight right right". 'Right' here of course, means right with respect to the direction of travel.

So both types of observed strategy reduce the majority of moves down to a choice of two: either straight and right, or (to avoid ambiguity), north and east. Why then is the latter one dominant? One possibility is that the **north and east** strategy is more robust. Obviously, to achieve a minimal path should be easier with the **rotational** strategy, but with both strategies it is necessary to make the right choices at the right time. If a wrong choice is made on the rotational

strategy then it can lead to a much longer path than if a wrong choice is made on the **north and east** strategy.

6.9.2 How did these Strategies Develop?

A commonly observed subsequence from earlier phases was of the sort $1 \rightarrow 3 \rightarrow 2$, so it is not possible to equate a vector principle as being necessarily linked to an adjacency principle. The progressive application of these two constraints simultaneously along with generalisation from what to do in one position to what to do in another, might lead to these sorts of strategy. The question is: can this modelling framework learn to apply these constraints?

Below are the constraints tried in the model which led to the identification of a serious difficulty.

Vectors Four additional output units were added to the output layer, corresponding to touching the stimulus in a position to the north, east, south and west of the position last touched. Note that these units do not directly support adjacent moves; if the last touch was in position 7 then the support for hitting a stimulus to the north would be added to the support for hitting the stimulus in location 4 *and* hitting the stimulus in location 1. Obviously, if these units do have an effect then the proportion of adjacent moves will increase. Each of these units was fully connected to the input layer and the activations and weights were allowed to have any positive or negative values, thus allowing each of these units to actively encourage or discourage moves in a particular direction.

Directions Directions are similar to vectors but relative to the direction of the 'path' through the grid. Only horizontal and vertical direction were considered. The direction of the path is given by the last two touches and is only defined if:

1. at least two touches have been made.

2. the last two touches were in the same row or the same column but not both.

If the path does not have a defined direction then these outputs have no effects. If a direction is defined, then there are four units that can add their support to the possible actions. These are:

Forwards hit a stimulus in the direction of the path.

Backwards hit a stimulus in the opposite direction of the path. Note that this does not necessarily cause a reiteration; for example, if the agent had touched a stimulus in position 1 followed by one in position 3, the support from the backwards unit would be added to position 1 *and* position 2.

Left hit a stimulus to the left of the path. e.g. if the path is north hit a stimulus to the west, if the path is east hit a stimulus to the north, etc.

Right hit a stimulus to the right of the path.

Rows and Columns Two output units one corresponding to: hit a stimulus in the same column, and the other to: hit a stimulus in the same row.

6.9.3 Summary of the Effect of the Added Constraints

The added actions changed the performance of the model such that qualitatively it looked much more similar to that of the monkeys. It was possible to look at the policy and see complex strategies. For example, with a certain four item trial the agent might behave thus: hit the stimulus in position 7 in the grid to begin with, then hit one to the north of it and adjacent, then move straight in the direction of travel and finally hit the stimulus in the centre. If this seems unnecessarily complicated, imagine someone giving directions to get to Edinburgh Castle from Bristo Square: "first go **across** the pedestrian crossing and head **straight** on, then go **north** along George IV Bridge, at the traffic lights **turn left**, and then keep walking **uphill**". The ability to incorporate different sorts of actions into an

action sequence allows highly complex strategies to be made up in the simplest possible way.

Unfortunately, although qualitatively the model with these actions did perform more similarly to the monkeys on the lower numbers, the model still failed to perform well on the higher number of stimuli. Rather than getting stuck in loops, as it did before, the model now just performs only very slightly better than chance. The reason for this was because all the strategies that were learned on the low numbers were unlearned on the high. Obviously, many of the early strategies were inappropriate for higher numbers but also, there came to light a serious problem for this type of model.

6.10 Forwards Errors

It was hoped that some 'fixed' sequence of actions would emerge. Suppose there was a fixed sequence such as $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$, then a forwards error occurs when a move such as $1 \rightarrow 3$ is made. The problem with the current method of evaluation is that this will be reinforced because the agent will believe it has found a shortcut. In more conventional sequence learning paradigms this is not a problem because the subject will get immediate feedback highlighting the mistake. For example, in the simultaneous chaining paradigm developed by Straub and Terrace(1981), when a subject makes a forwards error the trial is immediately terminated, no reward is given and there is a punishment delay before the next trial begins.

6.10.1 Identifying the Problem

At the moment the current learning algorithm is:

$$\epsilon_{t+1} = r_{t+1} + \gamma V_{t+1} - V_t$$

with the rationale being that $r_{t+1} + \gamma V_{t+1}$ and V_t are both predictions of the same thing but that the former is based on more up-to-date information. If a

forwards error occurs then V_{t+1} will be much higher than expected and the error will be strongly positive making the error more likely to occur again in the future. A large positive error is thus a signal that a forwards error has occurred, but there is also the possibility that a shortcut was actually found or that V_t greatly underestimated the future rewards. In an exhaustive search task there are no shortcuts by definition – but nobody informed the monkeys that the task was an exhaustive search task!

Forward Errors Lead to Unlearning

If a subject makes a forwards error they arrive at the end of the sequence without exhausting the set. The last transition $4- > 5$ will receive a strong negative reinforcement. As they do not know where the error occurred we can assume that they move randomly. If they hit the neglected stimulus then the least damage is done to the policy, but there has been a strong reinforcement to a transition that was not previously in the sequence. Wherever else they move to is treated as a backwards error so the move is less likely to happen again. If they do not move far enough back they will still miss the neglected item or items and the transitions upto the end will get negative reinforcement destabilising the sequence. If the move goes too far back the damage is not so great but of course this time round the same forwards error is even more likely to occur. If the correct transition is made then it will be reinforced because now the transition exhausts the set and a peanut is dispensed, but this self-repair is short lived because in future trials the transition will get negative reinforcement because it has now become overvalued.

6.10.2 A Possible Fix?

In order to overcome the problem of forwards errors it is necessary first to detect them and then to do something about it. It is impossible to do this for general tasks, but as in the exhaustive search task there are no shortcuts by definition, so it might be worthwhile to add a fix here for solving this task just so we can see if the constraints added can work given a chance.

In order to detect a forwards error it is necessary first to make a big assumption, namely that the evaluations are right. Then it is possible to define a forwards error as a transition that get a positive error greater than or equal to that which would have been received if a state two steps closer to the end had been reached.

If the evaluations are correct then

$$V_t = r_{t+1} + \gamma V_{t+1}$$

and

$$V_{t+1} = r_{t+2} + \gamma V_{t+2}$$

so

$$V_{t+2} = (V_{t+1} + r_{t+2})/\gamma$$

Substituting V_{t+2} in place of V_{t+1} in the error function we get, after a bit of cancellation:

$$\epsilon = r_{t+1} - r_{t+2} + V_{t+1} - V_t$$

Now the agent knows the values of V_t, V_{t+1}, and r_{t+1}, but it does not know the value of r_{t+2}. However, given that in this task the rewards for individual touches are mostly the same we might use the value of r_{t+1} as an estimate for r_{t+2}, which leads us to the definition of a forwards error as a transition for which the error is positive and bigger than the difference between V_{t+1} and V_t.

Having detected the error it is necessary to do something about it. The obvious and most simple thing to do is to reverse the sign of the error making it less likely to happen again. There are two problems with this; firstly significant damage is caused after the error has taken place but this could be overcome by turning off learning or something else similar. The second problem is the result of the assumption that the evaluations are initially correct. If the large positive error was not due to a forwards error but the result of an undervaluing of V_t, then it will get a large negative reinforcement which leads to even further undervaluing of V_t, which in turn leads to another false detection of a forwards error and so on.

It is difficult to see how to detect a forwards error with making assumptions about the correctness of the evaluations and this assumption is disastrous.

6.11 Dividing Sequences into Subsequences

If there is no easy way to detect forwards errors then perhaps there is a way to reduce the liklihood of them happening. The model works fine upto five stimuli, and forwards errors may occur but they do not occur frequently enough to cause major damage. As the number of stimuli increases the chance of making a forwards error towards the beginning of the sequence increases, and forwards errors at the beginning of the sequence are the most damaging. If there was a way of dividing bigger sequences into a number of shorter sequences then perhaps the problem of forwards errors could be avoided.

This is purely speculative. There are a number of difficult problems to be overcome before such an idea could be implemented in the model. However, there are hints in the monkeys behaviour that lend support to this idea. Firstly, in describing Alfie's behaviour I mentioned that he often cut across corners, missed out the centres of sides and went into the middle too soon. Perhaps it is the case that this was because Alfie was not particularly patient, but the fact that it happened so often and persisted for such a long time might be because they were undetected forwards errors and so were positively reinforced by Alfie's internal evaluator. Add to this the fact that Alfie was the only monkey to dramatically change his strategy, (and that forwards errors lead to unlearning), and things start falling into place. Considering Charlie's strategy, the characterisation as "north or east adjacently until you can't go any further, then jump" cannot, on its own, account for the high level of performance that Charlie reached. However, his high level of performance can be understood in terms of specific subsequences of these moves which all end at position 1. Charlie then might only need to remember to take the left path followed by the right path followed by the middle path, during a trial. If this was the case then why couldn't Charlie produce a strategy more

similar to the children. Well the argument still stands that he is trying to reduce the basic components of his strategy to a simplicity beyond that of the childrens'. Furthermore, it would seem that he likes his subsequences to start and finish in the same place.

As I said at the beginning of this section, this is all speculation, and it is difficult to test with the current experimental paradigm, as there is no fixed sequence that the subject must follow. Thus, we have to interpret both what is part of a sequence and what is an error. However, the idea of forwards errors leading to the unlearning of a strategy would explain why five year old children can learn to seriate five items "in a staircase", but not ten. Which then suggests that seven year olds not only succeed by using data-reduction strategies using the redundancy the task affords, but maybe also better abilities to chunk the task into smaller manageable pieces.

6.12 Summary

When this modelling started nobody knew just how well the monkeys would perform. In the end neither the monkeys nor the models reached perfect behaviour on the higher numbers of stimuli, although the monkeys came much closer. Both the monkeys and the models improved with experience and the performance of the model, given enough memory, was similar to monkeys up to five or six items. However the monkeys were able to improve further, whereas the models' performance deteriorated once the number of items increased to seven. When the task was analysed at the beginning it was noted how difficult it would be to evaluate performance and this proved to be correct. It seems to be impossible to have both a psychologically plausible representation of working memory and the ability to generate the correct evaluations of each touch at the same time. Having given up psychological plausibility, the models were unable to learn the correct evaluations. Increasing the potential power of the policy by elaborating it with new representational features and actions could not overcome the evaluation problem. Forwards errors are pathological to this type of self-regulation in an exhaustive

search task. The only way to overcome this problem, it seems, is to avoid it by chunking longer sequences into a number of shorter sequences. There are clues in the subjects' performance to support the hypotheses that they do make forwards errors, that these forwards errors lead to the unlearning of the strategy and that the more successful monkeys might be chunking, but it is impossible to test these with this paradigm.

6.12.1 Could Another Type of Model Fare Better?

These models used a connectionist instantiation of Sutton and Barto's Temporal Difference Learning framework and an incremental modelling strategy. Could, say a stack model, have done any better? It is important to remeber here what the objectives of the modelling were. We wanted to assess exactly what resources are required for this task, to understand how the agent could self-constrain its searches despite the minimal feedback, and to produce a model which includes as few assumptions as possible about the exact task it is doing, and which captures not only the final competence but also the trajectory to expertise. These requirements were beyond the modelling presented here but they are extremely tough requirements for any type of model to meet.

The reason these models failed was entirely due to the problem of forwards errors. This problem is not restricted to the specific learning algorithm used here, but is a very general problem. Any model that attempts to evaluate performance by estimating the distance to the goal state, will have to face this problem unless it has a completely different concept of state which somehow avoids the repeated states problem, and I cannot see how this could be done without building in the knowledge that the task is an exhaustive search task.

It is trivial to come up with a stack of rules which produce optimal searches every time, but this is entirely uninteresting. In other words, a model is not an alternative to this one unless it can capture the self-regulated constraints, and therefore the trajectory to expertise.

One might say with hindsight that the connectionist instantiation constrained the adding of resources in an unhelpful way, but considering that these resources must be used to produce soft mutual constraints it is difficult to think of a better implementational medium that connectionist networks.

I am neither committed to connectionist networks nor to Temporal Difference Learning. Both were chosen because the characteristics they possess were appropriate to modelling the task at hand in the way I wished to model it. What they have shown (yet again) is that induction on its own can only take you a small distance. What is needed for this task is some chunking device to reduce the inductive problem to a manageable proportion. An alternative model with chunking already included may well fare better than the ones here, but then these models with chunking added should do as well as any alternative.

Chapter 7

Summary and Conclusions

7.1 The Story So Far

The aim of this thesis was to model certain aspects of cognitive development. Our understanding of developmental processes is slight but growing. The modelling must be evaluated with respect to the wider research programme investigating cognitive growth.

The story started with Piaget who the first to see systems growth as the major problem that it is. The core component of Piaget's philosophy was that the growth of knowledge must be the result of an interactive process between an autonomous agent and its environment. A child does not simply impose their view of reality onto the world whether it fits or not, nor do they passively respond to environmental constraints. Rather, the child actively constructs their own version of reality. There are two crucial points about this knowledge construction process. Firstly, it is the result of a self-regulating process through which aspects of the world and existing knowledge are integrated together, and secondly the resulting knowledge is not necessarily "true", but is necessarily useful.

The main questions that must be answered are: what are the advantages conveyed by specific developments and what is the driving force behind them? To answer these questions it is first necessary to determine the ontological status of specific abilities. Piaget thought that rationality was based on logic, that the ability to make transitive inferences relied on logic, and thus the ability to make transitive inferences marks the end-point of cognitive development. If rationality

is based on logic then it also makes sense for the driving force behind development to be truth preservation.

Empirical studies have since shown that children as young as four, monkeys and pigeons can make such inferences, indicating that far from being the end-point of development, the skill is close to the lower ontological bounds of the system. The same studies indicate that the basis of the skill is not logic but some simple ordering ability. Furthermore, the idea of truth preservation driving the development of more and more powerful logics leads to the transition problem. So we need to restart from the beginning: revise our theories of rationality, determine a new ontology and ontogeny, and work out what else the driving force might be.

Given that the ability to make transitive inferences is so primitive, it is a good place to start this new programme. The questions to be answered are: why is transitive inference so primitive, and how does this fit in with our ideas of development? Imposing transitivity as a default assumption is a pragmatic thing to do. We need something on which to base our decisions, and ordering a set of choices is a simple way of achieving this. Also, given that transitive relations are ubiquitous, this imposition of an order is likely to be appropriate as well as useful. So one can consider ordering choices as constructing a workable version of reality: we impose an order because we need some basis on which to make a choice, and to get the best order the things we are ordering should have as much influence on the process as possible. However, this does not work for intransitive relations such as a circular relation. To cope with intransitive relations as well, the primitive ability underlying transitive inference must be more general than a simple ordering device. This does not mean that imposing transitivity as a default assumption is a bad thing to do, but it does suggest that this assumption should be defeasible. This is one possible solution to one of the major problems of system growth: how can a system be functional yet adaptive at the same time?

Given this characterisation of the ability to make transitive inferences, subjects doing transitivity tasks should be modelled as developing agents.

7.1.1 Transitive Inference Modelling

A model of subjects doing the transitive inference task was produced based on Sutton and Barto's Temporal Difference Learning Model and it was found that a weak stochastic transitive bias was inherent. This weak stochastic bias could be amplified by adding units with constant activations, dubbed **bias units**, to the representation. The strength of the transitive bias could be directly manipulated through the number of bias units. This showed just how easy it is to impose a transitive bias. All that is required is to elaborate the representation with anything; random noise would be sufficient. Thus we have a model which is consistent with the fact that transitive inference is an ontologically primitive ability.

The **Bias Model** could not explain the drop in performance on the triadic transfer phase. An alternative set of elaboratory units called **contiguity units** were added, which learned which stimuli were presented together. This model showed a strong transitive bias, and a drop in performance with the triads. Thus, we have a less primitive model but with similar scope to that observed in the subjects.

The **Contiguity Model** could also learn and generalise appropriately on a circular set of premises. The **Basic Model** and the **Bias Model** could learn a circular set of premises but could not generalise appropriately. So, all the models are reasonably flexible with respect to this intransitive relation, but the **Contiguity Model** can actually apply transitivity or circularity as a default assumption depending on the relation it is learning.

When the models were taught the premises for a transitive relation and then given the corresponding circular relation, the **Contiguity Model** showed that the transitive bias could be defeasible, but it could no longer generalise appropriately. The **Basic Model** didn't show a strong transitive bias to begin with, so it was relatively easy for it to accommodate the additional premise. The **Bias Model** could only learn the additional premise if the transitive bias was sufficiently weak. Therefore, the transitive bias was not defeasible.

The **Contiguity Model** is the only model that can account for the strong transitive bias, the drop in performance on the triads, shown by monkeys (McGonigle and Chalmers, 1984), and the defeasibility of the bias with respect to the circular relation, shown by pigeons (von Ferson et al, 1991).

The success of the **Contiguity Model** vindicated the view taken that performance of the transitive inference task is the result of a self-regulatory process involving specific generalisation devices that are applied to an indeterminate relation, making the learning of the relation easier and providing appropriate defeasible generalisations at the same time.

7.1.2 The New Ontology

Given that the ability to make transitive inferences relies on some simple ordering skill of limited scope, it would have been necessary to invent the seriation task if it had not existed already. Seriation is a much more transparent ordering skill, and the number of items to be seriated can be systematically increased to determine the scope of the underlying ability at different stages of development.

The seriation task did exist, however, invented by Piaget, but with a completely different ontology and ontogeny in mind. The ability to seriate items according to size was considered by Piaget to be a necessary precursor to the ability to make transitive inferences, but a child of four can reason transitively about a set of items without being able to seriate them. This seemingly paradoxical situation can be resolved only by understanding development as a process that increases the scope of certain primitive existing abilities. With this in mind, we must now attack the questions: what exactly is it that develops that increases the scope, what is the basic underlying ability, and what is the driving force behind the development? To answer these questions we must understand the basis of success, (and failure), on seriation tasks at different stages of development. The classical seriation task is not appropriate for this because it requires a number of different skills for success.

McGonigle and Chalmers decomposed the classical seriation task into four subtasks each of which isolated particular component skills that could affect com-

petence on the classical task. They found that the ability to calculate ordinal positions did not play an important role in seriation, and the form of seriation was the greatest determiner of success. The important development, it seems, is not some superior ordinal calculative ability, but specific data-reducing strategies that allow the subjects to reduce the task to a simple search task.

Thus, the picture of development changed considerably. We start life with a few basic abilities of limited scope. What develops is data-reducing strategies that allow us to solve more complicated tasks by reducing them such that they fall within the scope of one of the original basic abilities.

The questions of how these strategies develop and what is the driving force behind the process, still remain. This new characterisation suggests that resource limitations are the major constraints on our task-solving achievements, so it makes sense for resource management to be the driving force behind development.

This brings us to what might be termed the second major problem of systems growth: how can a system adapt whilst remaining adaptive? We are systems of limited cognitive resources. As we develop our behaviour becomes more constrained such that we can efficiently solve certain tasks, but at the cost of not being able to solve others. If this is all that happens, then sooner rather than later, we would become completely constrained and development would grind to a halt. However, human beings, at least, avoid or cope with this problem; we remain adaptive throughout our lives. To account for this we must accept that development not only increases our behavioural repertoire, but must also manage to deal with the increased complexity of our cognitive systems. Again, intelligent resource management must be the answer. We become more complex, and more constrained, but not in any way. We do it in the right way, the most efficient way. This is extremely vague, and the exhaustive search task was designed for the precise purpose of observing this process in action in order to elucidate what it is that is actually happening.

7.1.3 Exhaustive Search

The exhaustive search task requires subjects to touch every stimulus presented to them on a touchscreen before they receive any discriminative feedback. The subjects are given the freedom to touch each stimulus as many times and at any point in the production that they wish. However, repetitions and reiterations can only reduce the efficiency of the search. It is upto the subjects to self-determine what the best search strategy is – nobody told the monkeys or the children that it was an exhaustive search task. As the number of items to be searched increases the task becomes harder. It becomes harder because it becomes increasingly difficult for subjects to remember where they have touched and where they have not. The successful children avoid this problem of memory by constraining their searches, using adjacency and vector principles, such that it is easy to work out, at any point in the sequence, where they have been and where they have not. Monkeys begin by being inefficient in their searches but improve steadily with experience. As they improve, similar constraints to those which successful children use, emerge, but not to the same extent.

The aim of the modelling was to investigate what minimal cognitive resources might be required to achieve efficient search through self-supervised learning, and to see what constraints might emerge from this process or need to be explicitly applied in some way or other, to achieve efficient searches. When the number of items is less than six, then, with only a minimal amount of memory as to where the agent has been, the agent can learn to search efficiently fairly easily. The searches do not look similar to the monkey's or children's searches, indicating that adjacency and vector constraints do not emerge as automatically beneficial constraints from searching in such a constrained space as a 3x3 grid. In fact the only way to get the model to apply an adjacency constraint with a similar force to the subjects, was to bias the rewards slightly in favour of adjacent moves. Vector constraints could be learned through elaborating the actions. No constraints added to the model including the two mentioned previously, significantly helped performance on searches of over five or six items. The reason for this was the problem of forwards errors which appears to be incurable within this self-supervised learning

framework. I strongly suspect that the problem of forwards errors will be more general than this framework however.

So we have a model of search with limited scope which would fit in very well with the theory of development if it weren't for the fact that monkeys and children can obviously overcome this problem. "Overcome" is a more appropriate description than "solve". It is suggested that subjects overcome this problem by avoiding it, and that they avoid it by chunking their productions into subsequences all of which are short enough to ensure that forwards errors are not a major problem. Chunking is, of course, a form of data-reduction, but from this modelling attempt one would have to conclude that it is not a strategy that emerges from applying various constraints but the application of a dedicated data-reducing device. It would be extremely difficult to incorporated this device into this modelling framework.

The exhaustive search task was designed in order to observe how data-reducing strategies can develop through applying constraints leading to improved performance. The monkeys surprised everyone by their performance whereas the modelling encountered one problem after another. Firstly, the critic could not calculate the correct evaluations for an optimal policy with the plausible instantiation of memory due to representational constraints. Secondly, once this problem was resolved, the difficulty in learning a strategy proved significantly harder than anticipated. Thirdly, it was impossible to get the model to spontaneously make use of an adjacency constraint. And fourthly, the problem of forwards errors meant that the models could never cope with higher numbers of stimuli, so it was impossible to properly investigate the self-imposed application of constraints. Despite all this, the conclusions to be drawn from the modelling are remarkably consistent with the performance of the monkeys. Because of the problems with evaluation, it would be expected that the basic components of strategies might become over simple, and this is what we find with the monkeys. Forwards errors lead to unlearning, and Alfie showed forwards errors and his strategy was unlearned. To avoid forwards errors it is necessary to chunk long sequences into smaller subsequences, and there

was evidence that the monkeys did this as well. In conclusion, the models were spectacularly unsuccessful, but the modelling was highly informative.

7.2 From Transitive Choice to Full-Blown Logical Transitive Inference

The modeling of transitive inference presented in this thesis was based on the assumption that what had previously been considered as a task eliciting logical transitive inference was actually being solved by subjects using low-level transitive choice. It was then claimed that this low-level transitive choice was an in-built default assumption used to allow efficient decision-making over a specific set of objects. Although in general, we may not use high-level logical transitive inference in our everyday decision-making, this type of reasoning is perfectly valid and is used extensively by logicians and other mathematicians. This raises a question as to the relationship between the two types of transitive reasoning. It is possible that the higher type of reasoning is reliant on the concrete transitive choice mechanisms, but it is also possible that the two types of transitive reasoning are completely independent both from a computational and an ontological point of view.

One possible scenario is that in addition to the low-level transitive choice mechanisms, there is a metacognitive abstraction device that abstracts away from the specific items being reasoned about and classifies the different types of relations between specific objects, which eventually will produce some high-level logical transitive inference rule which works over abstract symbols. We cannot rule this out but, despite the intense interest over the relationship between subsymbolic and symbolic processing engendered by the rise of Connectionism, nobody has yet come up with a plausible account of how to make this transition work.

A more likely scenario is that the two are largely independent. Low-level transitive choice is a fundamental part of the cognitive architecture whereas high-level transitive inference is what Vygotsky would have referred to as a socioeconomic-cultural tool. Vygotsky provides a striking example of how socioeconomic cultural

change can affect high-level cognition in a naturally occurring experiment involving illiterate peasants working on small farms under a feudal lord in a remote area of the Soviet Union (Luria,1976). As part of a movement towards a socialist state, they became involved in collective farming practices that required numerous meetings to plan production and other matters. They also learned to read and write. Among the illiterate peasants' without these new experiences, classification, concept formation, reasoning and problem-solving skills were confined to the concrete and practical. For example, when told that all bears in the far north are white, the peasants would not predict the colour of a particular bear. A typical reply was "I don't know what colour the bears there are, I never saw them". After even minimal schooling the farm workers, in contrast, could consider this logical problem in the abstract and deduce the correct answer. Class inclusion problems are not transitive inference problems, but it would be highly surprising if we did not find a similar result with the latter.

Of course, this still does not specify the relation between concrete practical reasoning and abstract logical reasoning; saying that abstract logical reasoning is a cultural tool does not constrain the relationship in any way. If, on the other hand, we consider these cultural tools as cognitive strategies that increase the scope of pre-existing abilities, then it is plausible that, whilst concrete transitive choice is not a causal precursor to logical transitive inference, it may well be a necessary prerequisite.

7.3 Evaluating The Temporal Difference Learning Modelling Framework

It was never the intention of this thesis to evaluate the framework, just to use it as a tool to investigate developmental processes, but having used it extensively it is worthwhile noting down my reflections on it.

It has served its purpose well on the whole. As a framework for an exploratory modelling strategy it has proved considerably adaptable. It is possible to add

resources by elaborating the representation, and constraints can be applied in a number of different ways. Firstly, by elaborating the outputs with data-reducing actions, secondly through modifying the rewards, as used to capture adjacency, and thirdly by including specialised units with specialised evaluators such as with the starting points. I believe, though, that it has now reached the limits of its scope. It would be simple to add extra actions which represented a chunking; e.g. add an action that hit the stimuli in positions 7, 4, and 1, all in one go, and given the right chunks, this would certainly allow the model to perform as well as the children let alone the monkeys. However, these chunks do not naturally emerge from applying more plausible constraints and it seems as if what is required is some dedicated chunking device. It is difficult to see how this might be incorporated into the current framework. The problem is that, it looks as if, the changes required would be greater than what already exists.

7.3.1 Could the Framework be Extended?

The Temporal Difference Learning modelling framework could certainly be extended; in fact it already has, but in the wrong way for the wrong reasons. Sutton and Barto developed their **Temporal Difference Learning Model** from trying to model natural phenomena, i.e. animal conditioning. This idea of extending and developing learning models through modelling natural phenomena is the right way forward. Unfortunately, this is not what has happened.

With the version of the framework that I have used, it is impossible to get a convergence proof for the learning of a particular task. As a psychologist I do not expect such a thing, as I see dead-ends and sub-optimal behaviour in virtually everything I look at. Mathematicians, on the other hand find this messy and undesirable. Therefore Q-learning has been incorporated in order to obtain a convergence proof.

In the Q-learning versions of the model I have seen, there is no policy independent from the evaluations. Learning proceeds thus: evaluate the current state, and then work out the evaluations of all the possible states that could be reached

from this state and then take the best of them. Use this to alter the evaluation function. As this is used as a policy as well, (maybe some stochastic element is added), then applying it to the exhaustive search task would maximize the number of forwards errors!

These extended frameworks have been applied to search problems. However, these search problems were not designed to be given to subjects nor were they some abstraction of a naturally occurring problem for real agents. In fact it is very difficult to see why they were designed at all! In these tasks the agent is placed at some point in a simplified environment. The task is to find a goal at a particular location. Each trial, the starting point is chosen at random from all points in the environment, but the goal remains in exactly the same place. From the ecological point of view this is completely back to front. Real agents do search for 'goals', but they tend to start from the same point each time, (their home), and the goals inconsiderately keep moving around!

If, on the other hand, the mathematicians and computer scientists would listen to psychologists, then I have no doubt that this modelling framework could be extended in a highly productive way.

7.4 Classification and Chunking

Classification and chunking have been studied for a long time, but the research presented here provides a new perspective. It has always been obvious that we classify and chunk – language gives a most vivid example. It has also been obvious that these processes are data-reducing – that is the very essence of them. But the significance of this fact may not have been fully appreciated until now. Now – we are claiming that data-reducing strategies are the fundamental component of what it is that develops. Not only have we emphasized the importance, but *we have also better specified the purpose*: the purpose is to increase the scope of some basic ability such as search.

The modelling suggested that chunking was necessary to overcome the problem of forwards errors. However, 'chunks' did not emerge naturally from applying other constraints which were data-reducing in their own way. It was also difficult to see how to incorporate chunking in a plausible way into the modelling framework. As was stated, perhaps what is needed is some dedicated chunking device. The ontological status of this device must therefore be as a metacognitive device dedicated to development. Obviously, this type of device must become central to our investigations.

7.4.1 Object-Oriented Classification

It is easy enough to classify the world or chunk ones actions, but there are an infinite number of possible classifications. The difficult part is to determine the 'right' classification.

There are a number of similarities between the area of cognitive development and object-oriented programming (OOP); perhaps this area can inform us. OOP was designed specifically to cope with the problems associated with extending software systems – (the problem of systems growth again). One major problem of systems growth with software as with autonomous agents, is how to add something new without completely redesigning what you already have. The main component of the **object model** is abstraction which is just the other side of the coin from classification. Unfortunately, the **object model** does not specify which classification to use, it is left to the designer. The only advice you will find in any book on object-oriented design is to try a classificatory scheme and see if it works. If it doesn't then refine it. In other words, developing the 'right' classification should be an interactive process, with the criterion for success being whether it works or not. This is exactly what Piaget would have prescribed. Needless to say, considering we are talking about computers, hardware limitations can play a significant factor in determining the 'right' classification.

So, OOP was designed so that extendibility could be built in, and the means to do this is to have an easily refinable classificatory scheme. The actual process

of refining the classification must be interactive and the criterion of success is whether it works.

7.4.2 SOAR and Chunking

SOAR is a comprehensive cognitive architecture in the GPS mould, which Newell(1992) has put forward as a candidate unified theory of cognition. Its interest to us here is the central role that chunking plays within the system.

In SOAR all behaviour can be considered as a result of *searches* of the **problem space**. If the goal does not lie within the problem space the system can always posit a larger space, so theoretically any problem can be solved. Note that resource limitations play no role in SOAR whatsoever. This is true for memory as well. Long-term memory is a collection of production rules and during the **decision process** the actions coming from matching production rules are placed into working memory. There is no conflict resolution of the production rules so all possible actions are generated. These actions are in the form of preferences and once loaded into working memory they can also be used to match further production rules. This constitutes the **elaboration phase**. Once no more actions enter working memory the elaboration phase is said to have run to **quiescence**. The decision process then runs which simply implements the semantics of preferences. There are five possible outcomes to this procedure. Firstly, an unequivocal result might be produced in which case the action is simply carried out. Otherwise an **impasse** occurs. There are four possible impasses:

1. tie impasse – the collection of alternatives cannot be discriminated.

2. no-change impasse – there are no choices available.

3. reject impasse – the only preference may be for rejecting one of the decisions already made.

4. conflict impasse – two or more preferences may contradict each other.

Each time an impasse occurs, a subgoal is created. The knowledge of how to do this is given by the programmer. It is the formation and solution of subgoals that leads to chunking. When an impasse arises, the contents of working memory are stored and the subgoal is set up. Once the subgoal is achieved and the impasse resolved, new productions can be added to long-term memory with the contents of working memory that were used to achieve the subgoal as the head of the production rule and the actions taken at the end of the subgoal as the actions of the production rule. Thus, the information from the subgoal is chunked together into a new production which means that the impasse that occurred should not happen again. There will also be some enforced generalisation because not all the items in working memory are included in the head of the new rule, only those which contributed to the solution of the subgoal. Thus chunking within SOAR can be seen as the permanent cacheing of goal results. This leads to more efficient problem solving in the future.

It is difficult to know how SOAR would cope with the exhaustive search task, because a lot would depend on how it was set up – what initial knowledge would it need, both in terms of initial productions and knowledge of how to create an appropriate subgoal. Presumably the subgoal must be simpler than the goal which caused the impasse. The problem is how the knowledge of creating subgoals could be implemented without giving the agent knowledge of what the task is. The task given to the subjects is nothing more than to maximize the rate of return of peanuts. This task description must then be translated into a top level goal involving (I presume), touches to stimuli on the screen. If an impasse occurs in trying to solve this top-level goal, then a subgoal will be created which raises the opportunity for chunking. The chunks must involve more than one touch to the screen or else chunks are merely forwards errors.

SOAR does not give you any measure of how good its performance is – this must be included by the programmer. In fact, it seems as if SOAR was designed to think in terms of success and failure rather than better or worse performance. It would be exceedingly difficult, if this is true, to capture the level of performance of the monkeys and the improvement over time that they show. I expect that

SOAR would either reach optimal performance very quickly, or it would settle for longer runs and not improve at all, because it sees this runs as successful.

It is very difficult to imagine SOAR going through a hierarchy of subgoals without the whole sequence being constrained from the end. In other words the top-level goal must be "touch the stimulus that gives the reward" but because of an impasse it must set up the subgoal "touch the stimulus that must be touched immediately before the touch that gives a reward" and so on. There is strong empirical evidence that this is not the way subjects learn series. In the exhaustive search task there was strong evidence for preferred starting points but no evidence whatsoever, for preferred end-points. Straub et al(1979), managed to train pigeon to peck sequences of colours but only if given in a forwards training schedule, ie $A, AB, ABC, ABCD$. If the pigeons were given the backwards schedule, $D, CD, BCD, ABCD$, they continued to persist in pecking D first for upto 25 days! I do not see how SOAR could capture this sort of data without allowing the way in which a subgoal was achieved to have an effect of the supergoal which created it. This may be allowed in SOAR but it seems to go totally against the whole ethos of the top-down problem-solving approach on which SOAR is based.

Finally, chunking in SOAR seems to have been included to improve the efficiency of problem-solving, but it does not have any effect of the actual performance of the model, because the resources that the model can use are unlimited. What is suggested by the modelling presented in this thesis, is just the opposite – that chunking is needed not for efficiency but for adequacy, and that throwing more resources at the model does not make any difference.

7.4.3 The Empirical Programme Proceeds

Of course, the empirical programme has proceeded from where I left it, and it has proceeded to look exactly at the ability to chunk and classify. Subjects are presented a number of different looking stimuli on the screen and are required to touch them in a specific order. Forcing the subjects to touch stimuli in a specific order means that they can no longer use the spacial strategies that they developed

for the exhaustive search task. Once the subjects have learned a sequence such as $A \rightarrow B \rightarrow C$, they are given transfer tests in which each of the nodes can be overloaded such that, for example, there are two As, two Bs and two Cs. The questions are: to what extent can they transfer there knowledge of $A \rightarrow B \rightarrow C$ to $A \rightarrow A \rightarrow B \rightarrow B \rightarrow C \rightarrow C$? Does this transfer improve with experience? Is the maximum length of production affected by the ease with which the sequence can be chunked? What sort of errors are made? Are some types of classification easier to make than others? And are monkeys capable of hierarchical classification which must be an essential component of language in humans?

The experiment is still ongoing but the initial results are extremely promising. Five year old children show complete transfer, and when the nodes are overloaded their reaction times show strong phrasing effects. The monkeys transfer is by no means complete but is getting better and the phrasing effects are beginning to emerge. The production length of the monkeys is currently nine, ($AAABBBCCC$), which is longer than any sequence previously reported for a non-human subject. One subject, Ollie, has reached criterion on nine, (the equivalent of 75% minimal paths when translated into the exhaustive search task measure of success), and amazingly she succeeded on her first trial of $AAAABBBBCCCC$. Interestingly, in relation to the modelling, almost all the monkeys' errors are forwards errors.

Analysing the subjects' performance in detail, on a task where their chunking is totally transparent, should provide a wealth of clues as to the underlying mechanisms.

7.5 Final Conclusions

There are four basic principles concerning development that come out of this thesis. These principles do not come out of the modelling alone, but out of the research programme of which this modelling was a part. The role of the modelling has been to illustrate these principles, to identify some of the problems and to specify some of the details.

1. We have a finite set of basic abilities whose scope is great in terms of potentially applicability, but which are initially poor in both scope and efficiency.

2. The purpose of development is to increase the scope and efficiency of these basic abilities.

3. The driving force behind development cannot only be to optimize rewards from the environment, but must include optimizing the use of cognitive resources.

4. This resource management is manifested in terms of constraining behaviour appropriately and producing data-reducing strategies.

"Resource management" and "data-reducing strategy" are not well-defined concepts at the present time. We have a good general idea about what they mean: resource management refers to an active self-regulatory process whose purpose is to make optimal use of a limited set of resources, but we do not exactly what these resources are, nor how the self-regulation works. A data-reducing strategy is a strategy that transforms a complex task which we cannot solve into a simpler task that we can. For example, we may have problems solving a task with nine entities, and a data-reducing strategy might transform this task into three subtasks with three entities each, which we can solve very easily. However, we do not know how they are represented, we are ignorant about the range of basic abilities and their scope for which the strategies must work, and we are unclear as to how

exactly they develop. However, the foundations of a new theory of cognitive development are swiftly falling into place, and the research is becoming better and better focussed.

There is no other coherent theory of cognitive development to compete with the one emerging here, and since people became disillusioned with Piaget's theory, there has been a theoretical vacuum. I hope this theory will fill that gap. Afterall, we will not get a unified psychology until we have sorted out the problems of development.

Bibliography

Boden, M. (1979) *Piaget.* London: Fontana.

Bowlby, J. (1969) *Attachment and Loss*, Vol. 1: *Attachment.* London: Hogarth Press.

Breslow, L. (1981) Re-evaluation of the literature on the development of transitive inference. *Psychological Bulletin* **89**, 325–351.

Brofenbrenner, U., R. K. Silbereisen, K. Eyferth and G. Rudinger (1986) Recent advances in research on human development. In *Development as Action in Context: Problem Behaviour and Normal Youth Development.* New York: Springer-Verlag.

Brooks, R. (1991) Intelligence without reason. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, pp. 569–595.

Bryant, P. E. and T. Trabasso (1971) Transitive inference and memory in young children. *Nature* **232**, 456–458.

Chalmers, M. A. and B. O. McGonigle (1984) Are children any more logical than monkeys on the five-term series problem? *Journal of Experimental Child Psychology* **37**, 355–377.

Cummins (1983) *The nature of psychological explanation.* MIT Press.

De Soto, C. B., M. London and S. Handel (1965) Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology* **2**, 513–521.

Donaldson, M. (1978) *Children's Minds.* London: Fontana.

Edwards, W. (1954) The theory of decision making. *Psychological Bulletin* **51**, 380–417.

von Ferson, L., C. D. L. Wynne, J. D. Delius and J. E. R. Staddon (1991) Transitive inference formation in pigeons. *Journal of Experimental Psychology (Animal Behaviour Processes)* **17**(3), 334–341.

Freud, S. (1964) *New Introductory Lectures on Psycho-Analysis and Other Works: 1932–1936.* London: Hogarth Press and the Institute of Psycho-Analysis. Vol. 22 of complete psychological works, gen. ed. J. Strachey.

Gillan, D. J. (1981) Reasoning in the chimpanzee, II: transitive inference. *Journal of Experimental Psychology (Animal Behaviour Processes)* **7**, 150–164.

Gottlieb, G. and E. Hearst (1979) Comparative psychology and ethology. In E. Hearst, ed., *The First Century of Experimental Psychology.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Guiselin, M. T. and F. M. Scudo (1986) The bioeconomics of phenotype selection. *Behavioural and Brain Sciences* **9**, 194–195. Comment on D. Vining, 'Social versus reproductive success: The central theoretical problems of human sociobiology'.

Harris, M. R. (1988) *The Computational Modelling of Transitive Inference.* PhD thesis, Department of Artificial Intelligence, University of Edinburgh.

Harris, M. R. (in press) A model of transitive choice. *Quarterly Journal of Experimental Psychology.*

Hicks, V. C. and H. A. Carr (1912) Human reactions in a maze. *Journal of Animal Psychology* **2**, 98–125.

Huttonlocher, J. (1968) Constructing spatial images: A strategy in reasoning. *Psychological Review* **75**, 550–560.

Inhelder, B. and J. Piaget (1964) *The early growth of logic in the child.* London: Routledge and Kegan Paul.

Johnson-Laird, P. N. (1983) *Mental Models.* Cambridge: Cambridge University Press.

Kallio, K. D. (1982) Developmental change on a five-term transitive inference. *Journal of Experimental Child Psychology* **33**, 142–164.

LeGare, M. (1987) The use of General Systems Theory as a metatheory for developing and evaluating theories in the neurosciences. *Behavioral Science* **32**, 106–120.

Lorenz, K. Z. (1937) Uber die Bildung des Instinktbegriffes. *Die Naturwissenschaften* **25**, 289–300, 307–318, 325–331.

de Lorenzana, J. M. A. and L. M. Ward (1987) On evolutionary systems. *Behavioral Science* **32**, 19–33.

Lunzer, E. A. and R. Lucas (1977) When do children acquire transitive inferences? Unpublished ms.

McGonigle, B. O. (1987) Nonverbal thinking by animals? *Nature* **325**, 110–112.

McGonigle, B. O. and M. A. Chalmers (1977) Are monkeys logical? *Nature* **267**, 694–696.

McGonigle, B. O. and M. A. Chalmers (1986) Representations and strategies during inference. In T. F. Myers, E. K. Brown and B. O. McGonigle, eds., *Reasoning and Discourse Processes.* London: Academic Press.

McGonigle, B. O. and M. A. Chalmers (1992) Monkeys are rational! *Quarterly Journal of Experimental Psychology* **45B**(3), 189–228.

McGonigle, B. O. and M. A. Chalmers (in press) *Intelligent Systems: A cognitive analysis.* New York: Columbia University Press.

Minsky, M. and S. Papert (1969) *Perceptrons: An Introduction to Computational Geometry.* Cambridge, Mass.: MIT Press.

Munn, N. L. and L. Carmichael (1954) Learning in children. In L. Carmichael, ed., *Manual of Child Psychology*, 2nd ed. New York: John Wiley and Sons.

Neapolitan, D. M. (1991) *A Micro-Analysis of Seriation Skills.* PhD thesis, Centre for Cognitive Science, University of Edinburgh.

Newell, A., J. C. Shaw and H. A. Simon (1958) Elements of a theory of human problem solving. *Psychological Review* **65**, 151–166.

Piaget, J. and B. Inhelder (1973) *Memory and Intelligence.* London: Routledge and Kegan Paul.

Reder, L. M. and J. R. Anderson (1980) A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology* **12**, 447–472.

Rumelhart, D. E., G. E. Hinton and R. J. Williams (1986) Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, eds., *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: *Foundations*, pp. 318–362. Cambridge, Mass.: MIT Press.

Sherry, D. F. and D. L. Schacter (1987) The evolution of multiple memory systems. *Psychological Review* **94**(4), 439–454.

Simon, H. A. and A. Newell (1963) The uses and limitations of models. In *Theories of contemporary psychology.*. New York: MacMillan.

Skinner, B. F. (1948) *Walden Two.* New York: Macmillan.

Smedland, J. (1966) The development of concrete transitivity of length in children. *Scandinavian Journal of Psychology* **7**, 81–92.

Smith, E. E., N. Adams and D. Schorr (1978) Fact retrieval and the paradox of interference. *Cognitive Psychology* **10**, 438–464.

St Johnston, B. (1989) Towards a connectionist model of the development of transitive inference. Master's thesis, Centre for Cognitive Science, University of Edinburgh.

Straub, R. O., M. S. Seidenberg, T. G. Bever and H. S. Terrace (1979) Serial learning in the pigeon. *Journal of the Experimental Analysis of Behavior* **32**(2), 137–148.

Straub, R. O. and H. S. Terrace (1981) Generalisation of serial learning in the pigeon. *Animal Learning and Behaviour* **9**(4), 454–468.

Sutton, R. S. and B. Pinette (1985) The learning of world models by connectionist networks. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society.* Distributed by Lawrence Erlbaum Associates, Hillsdale, N.J.

Sutton, R. S. and A. G. Barto (1987) A temporal-difference model of classical conditioning. Unpublished ms., March 1987.

Sutton, R. S. and A. G. Barto (1989) Time-derivative models of Pavlovian reinforcement. In *Learning and Computational Neuroscience.* Cambridge, Mass.: MIT Press.

Trabasso, T. and C. A. Riley (1975) On the construction and use of representations involving linear order. In *Information Processing and Cognition: The Loyola Symposium.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Tversky, A. (1969) Intransitivity of preferences. *Psychological Review* **76**(1), 31–48.

Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, Mass.: Harvard University Press.

Wilson, E. O. (1975) *Sociobiology: The New Synthesis.* Cambridge, Mass.: Harvard University Press.